

SEQUENCE LISTING

(1) GENERAL INFORMATION:

(i) APPLICANT: Bell, Graeme I.
Yamagata, Kazuya
Oda, Naohisha
Kaisaki, Pamela J.
Furuta, Hiroto
Horikawa, Yukio
Menzel, Stephen

(ii) TITLE OF INVENTION: MUTATIONS IN THE DIABETES SUSCEPTIBILITY
GENES HEPATOCYTE NUCLEAR FACTOR (HNF) 1 ALPHA, HNF-1BETA
AND HNF-4ALPHA

(iii) NUMBER OF SEQUENCES: 146

(iv) CORRESPONDENCE ADDRESS:

(A) ADDRESSEE: Arnold, White & Durkee
(B) STREET: P.O. Box 4433
(C) CITY: Houston
(D) STATE: Texas
(E) COUNTRY: USA
(F) ZIP: 77210

(v) COMPUTER READABLE FORM:

(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30

(vi) CURRENT APPLICATION DATA:

(A) APPLICATION NUMBER: US Unknown
(B) FILING DATE: Concurrently Herewith
(C) CLASSIFICATION: Unknown

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/029,679
(B) FILING DATE: 30-OCT-1996

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/028,056
(B) FILING DATE: 02-OCT-1996

(vii) PRIOR APPLICATION DATA:

(A) APPLICATION NUMBER: US 60/025,719
(B) FILING DATE: 10-SEP-1996

(viii) ATTORNEY/AGENT INFORMATION:

(A) NAME: Wilson, Mark B.
(B) REGISTRATION NUMBER: 37,259
(C) REFERENCE/DOCKET NUMBER: ARCD:272

(ix) TELECOMMUNICATION INFORMATION:

(A) TELEPHONE: 512/418-3000
(B) TELEFAX: 512/474-7577

(2) INFORMATION FOR SEQ ID NO:1:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3238 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base..
- (B) LOCATION: 988
- (D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A, C, G, or T"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(24..986, 990..1916)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

```
CGTGGCCCTG TGGCAGCCGA GCC ATG GTT TCT AAA CTG AGC CAG CTG CAG      50
      Met Val Ser Lys Leu Ser Gln Leu Gln
      1              5

ACG GAG CTC CTG GCG GCC CTG CTC GAG TCA GGG CTG AGC AAA GAG GCA      98
Thr Glu Leu Leu Ala Ala Leu Leu Glu Ser Gly Leu Ser Lys Glu Ala
10              15              20              25

CTG ATC CAG GCA CTG GGT GAG CCG GGG CCC TAC CTC CTG GCT GGA GAA     146
Leu Ile Gln Ala Leu Gly Glu Pro Gly Pro Tyr Leu Leu Ala Gly Glu
30              35              40

GGC CCC CTG GAC AAG GGG GAG TCC TGC GGC GGC GGT CGA GGG GAG CTG     194
Gly Pro Leu Asp Lys Gly Glu Ser Cys Gly Gly Gly Arg Gly Glu Leu
35              45              50              55

GCT GAG CTG CCC AAT GGG CTG GGG GAG ACT CGG GGC TCC GAG GAC GAG     242
Ala Glu Leu Pro Asn Gly Leu Gly Glu Thr Arg Gly Ser Glu Asp Glu
60              65              70

ACG GAC GAC GAT GGG GAA GAC TTC ACG CCA CCC ATC CTC AAA GAG CTG     290
Thr Asp Asp Asp Gly Glu Asp Phe Thr Pro Pro Ile Leu Lys Glu Leu
75              80              85

GAG AAC CTC AGC CCT GAG GAG GCG GCC CAC CAG AAA GCC GTG GTG GAG     338
Glu Asn Leu Ser Pro Glu Glu Ala Ala His Gln Lys Ala Val Val Glu
90              95              100              105

ACC CTT CTG CAG GAG GAC CCG TGG CGT GTG GCG AAG ATG GTC AAG TCC     386
Thr Leu Leu Gln Glu Asp Pro Trp Arg Val Ala Lys Met Val Lys Ser
110              115              120

TAC CTG CAG CAG CAC AAC ATC CCA CAG CGG GAG GTG GTC GAT ACC ACT     434
Tyr Leu Gln Gln His Asn Ile Pro Gln Arg Glu Val Val Asp Thr Thr
125              130              135

GGC CTC AAC CAG TCC CAC CTG TCC CAA CAC CTC AAC AAG GGC ACT CCC     482
Gly Leu Asn Gln Ser His Leu Ser Gln His Leu Asn Lys Gly Thr Pro
140              145              150

ATG AAG ACG CAG AAG CGG GCC GCC CTG TAC ACC TGG TAC GTC CGC AAG     530
```

	Met	Lys	Thr	Gln	Lys	Arg	Ala	Ala	Leu	Tyr	Thr	Trp	Tyr	Val	Arg	Lys	
	155						160					165					
5	CAG	CGA	GAG	GTG	GCG	CAG	CAG	TTC	ACC	CAT	GCA	GGG	CAG	GGA	GGG	CTG	578
	Gln	Arg	Glu	Val	Ala	Gln	Gln	Phe	Thr	His	Ala	Gly	Gln	Gly	Gly	Leu	
	170					175					180					185	
10	ATT	GAA	GAG	CCC	ACA	GGT	GAT	GAG	CTA	CCA	ACC	AAG	AAG	GGG	CGG	AGG	626
	Ile	Glu	Glu	Pro	Thr	Gly	Asp	Glu	Leu	Pro	Thr	Lys	Lys	Gly	Arg	Arg	
					190					195					200		
15	AAC	CGT	TTC	AAG	TGG	GGC	CCA	GCA	TCC	CAG	CAG	ATC	CTG	TTC	CAG	GCC	674
	Asn	Arg	Phe	Lys	Trp	Gly	Pro	Ala	Ser	Gln	Gln	Ile	Leu	Phe	Gln	Ala	
				205					210					215			
20	TAT	GAG	AGG	CAG	AAG	AAC	CCT	AGC	AAG	GAG	GAG	CGA	GAG	ACG	CTA	GTG	722
	Tyr	Glu	Arg	Gln	Lys	Asn	Pro	Ser	Lys	Glu	Glu	Arg	Glu	Thr	Leu	Val	
			220					225					230				
25	GAG	GAG	TGC	AAT	AGG	GCG	GAA	TGC	ATC	CAG	AGA	GGG	GTG	TCC	CCA	TCA	770
	Glu	Glu	Cys	Asn	Arg	Ala	Glu	Cys	Ile	Gln	Arg	Gly	Val	Ser	Pro	Ser	
			235				240					245					
30	CAG	GCA	CAG	GGG	CTG	GGC	TCC	AAC	CTC	GTC	ACG	GAG	GTG	CGT	GTC	TAC	818
	Gln	Ala	Gln	Gly	Leu	Gly	Ser	Asn	Leu	Val	Thr	Glu	Val	Arg	Val	Tyr	
	250					255					260					265	
35	AAC	TGG	TTT	GCC	AAC	CGG	CGC	AAA	GAA	GAA	GCC	TTC	CGG	CAC	AAG	CTG	866
	Asn	Trp	Phe	Ala	Asn	Arg	Arg	Lys	Glu	Glu	Ala	Phe	Arg	His	Lys	Leu	
					270					275					280		
40	GCC	ATG	GAC	ACG	TAC	AGC	GGG	CCC	CCC	CCA	GGG	CCA	GGC	CCG	GGA	CCT	914
	Ala	Met	Asp	Thr	Tyr	Ser	Gly	Pro	Pro	Pro	Gly	Pro	Gly	Pro	Gly	Pro	
				285					290					295			
45	GCG	CTG	CCC	GCT	CAC	AGC	TCC	CCT	GGC	CTG	CCT	CCA	CCT	GCC	CTC	TCC	962
	Ala	Leu	Pro	Ala	His	Ser	Ser	Pro	Gly	Leu	Pro	Pro	Pro	Ala	Leu	Ser	
			300					305					310				
50	CCC	AGT	AAG	GTC	CAC	GGT	GTG	CGC	TNT	GGA	CAG	CCT	GCG	ACC	AGT	GAG	1010
	Pro	Ser	Lys	Val	His	Gly	Val	Arg		Gly	Gln	Pro	Ala	Thr	Ser	Glu	
			315				320						325				
55	ACT	GCA	GAA	GTA	CCC	TCA	AGC	AGC	GGC	GGT	CCC	TTA	GTG	ACA	GTG	TCT	1058
	Thr	Ala	Glu	Val	Pro	Ser	Ser	Ser	Gly	Gly	Pro	Leu	Val	Thr	Val	Ser	
			330				335					340					
60	ACA	CCC	CTC	CAC	CAA	GTG	TCC	CCC	ACG	GGC	CTG	GAG	CCC	AGC	CAC	AGC	1106
	Thr	Pro	Leu	His	Gln	Val	Ser	Pro	Thr	Gly	Leu	Glu	Pro	Ser	His	Ser	
			345			350											

		395		400		405												
5		GGG GTC ATG ACC ATC GGG CCT GGT GAG CCT GCC TCC CTG GGT CCT ACG		1298														
		Gly Val Met Thr Ile Gly Pro Gly Glu Pro Ala Ser Leu Gly Pro Thr																
		410		415		420												
10		TTC ACC AAC ACA GGT GCC TCC ACC CTG GTC ATC GGC CTG GCC TCC ACG		1346														
		Phe Thr Asn Thr Gly Ala Ser Thr Leu Val Ile Gly Leu Ala Ser Thr																
		425		430		435												
		CAG GCA CAG AGT GTG CCG GTC ATC AAC AGC ATG GGC AGC AGC CTG ACC		1394														
		Gln Ala Gln Ser Val Pro Val Ile Asn Ser Met Gly Ser Ser Leu Thr																
		445		450		455												
15		ACC CTG CAG CCC GTC CAG TTC TCC CAG CCG CTG CAC CCC TCC TAC CAG		1442														
		Thr Leu Gln Pro Val Gln Phe Ser Gln Pro Leu His Pro Ser Tyr Gln																
		460		465		470												
20		CAG CCG CTC ATG CCA CCT GTG CAG AGC CAT GTG ACC CAG AGC CCC TTC		1490														
		Gln Pro Leu Met Pro Pro Val Gln Ser His Val Thr Gln Ser Pro Phe																
		475		480		485												
25		ATG GCC ACC ATG GCT CAG CTG CAG AGC CCC CAC GCC CTC TAC AGC CAC		1538														
		Met Ala Thr Met Ala Gln Leu Gln Ser Pro His Ala Leu Tyr Ser His																
		490		495		500												
30		AAG CCC GAG GTG GCC CAG TAC ACC CAC ACG GGC CTG CTC CCG CAG ACT		1586														
		Lys Pro Glu Val Ala Gln Tyr Thr His Thr Gly Leu Leu Pro Gln Thr																
		505		510		515												
35		ATG CTC ATC ACC GAC ACC ACC AAC CTG AGC GCC CTG GCC AGC CTC ACG		1634														
		Met Leu Ile Thr Asp Thr Thr Asn Leu Ser Ala Leu Ala Ser Leu Thr																
		525		530		535												
40		CCC ACC AAG CAG GTC TTC ACC TCA GAC ACT GAG GCC TCC AGT GAG TCC		1682														
		Pro Thr Lys Gln Val Phe Thr Ser Asp Thr Glu Ala Ser Ser Glu Ser																
		540		545		550												
45		GGG CTT CAC ACG CCG GCA TCT CAG GCC ACC ACC CTC CAC GTC CCC AGC		1730														
		Gly Leu His Thr Pro Ala Ser Gln Ala Thr Thr Leu His Val Pro Ser																
		555		560		565												
50		CAG GAC CCT GCC GGC ATC CAG CAC CTG CAG CCG GCC CAC CGG CTC AGC		1778														
		Gln Asp Pro Ala Gly Ile Gln His Leu Gln Pro Ala His Arg Leu Ser																
		570		575		580												
55		GCC AGC CCC ACA GTG TCC TCC AGC AGC CTG GTG CTG TAC CAG AGC TCA		1826														
		Ala Ser Pro Thr Val Ser Ser Ser Ser Leu Val Leu Tyr Gln Ser Ser																
		585		590		595												
60		GAC TCC AGC AAT GGC CAG AGC CAC CTG CTG CCA TCC AAC CAC AGC GTC		1874														
		Asp Ser Ser Asn Gly Gln Ser His Leu Leu Pro Ser Asn His Ser Val																
		605		610		615												
65		ATC GAG ACC TTC ATC TCC ACC CAG ATG GCC TCT TCC TCC CAG		1916														
		Ile Glu Thr Phe Ile Ser Thr Gln Met Ala Ser Ser Ser Gln																
		620		625		630												
70		TAACCACGGC ACCTGGGCCC TGGGGCCTGT ACTGCCTGCT TGGGGGGTGA TGAGGGCAGC		1976														
		AGCCAGCCCT GCCTGGAGGA CCTGAGCCTG CCGAGCAACC GTGGCCCTTC CTGGACAGCT		2036														

5
 10
 15
 20
 25
 30
 35
 40
 45
 50
 55
 60

	Pro	Gly	Pro	Tyr	Leu	Leu	Ala	Gly	Glu	Gly	Pro	Leu	Asp	Lys	Gly	Glu	
			35					40					45				
5	Ser	Cys	Gly	Gly	Gly	Arg	Gly	Glu	Leu	Ala	Glu	Leu	Pro	Asn	Gly	Leu	
		50					55					60					
	Gly	Glu	Thr	Arg	Gly	Ser	Glu	Asp	Glu	Thr	Asp	Asp	Asp	Gly	Glu	Asp	
	65					70					75					80	
10	Phe	Thr	Pro	Pro	Ile	Leu	Lys	Glu	Leu	Glu	Asn	Leu	Ser	Pro	Glu	Glu	
					85					90					95		
	Ala	Ala	His	Gln	Lys	Ala	Val	Val	Glu	Thr	Leu	Leu	Gln	Glu	Asp	Pro	
15				100					105					110			
	Trp	Arg	Val	Ala	Lys	Met	Val	Lys	Ser	Tyr	Leu	Gln	Gln	His	Asn	Ile	
		115						120					125				
20	Pro	Gln	Arg	Glu	Val	Val	Asp	Thr	Thr	Gly	Leu	Asn	Gln	Ser	His	Leu	
		130					135					140					
	Ser	Gln	His	Leu	Asn	Lys	Gly	Thr	Pro	Met	Lys	Thr	Gln	Lys	Arg	Ala	
	145					150					155					160	
25	Ala	Leu	Tyr	Thr	Trp	Tyr	Val	Arg	Lys	Gln	Arg	Glu	Val	Ala	Gln	Gln	
				165						170					175		
	Phe	Thr	His	Ala	Gly	Gln	Gly	Gly	Leu	Ile	Glu	Glu	Pro	Thr	Gly	Asp	
30				180					185					190			
	Glu	Leu	Pro	Thr	Lys	Lys	Gly	Arg	Arg	Asn	Arg	Phe	Lys	Trp	Gly	Pro	
		195						200					205				
35	Ala	Ser	Gln	Gln	Ile	Leu	Phe	Gln	Ala	Tyr	Glu	Arg	Gln	Lys	Asn	Pro	
		210					215					220					
	Ser	Lys	Glu	Glu	Arg	Glu	Thr	Leu	Val	Glu	Glu	Cys	Asn	Arg	Ala	Glu	
	225					230					235					240	
40	Cys	Ile	Gln	Arg	Gly	Val	Ser	Pro	Ser	Gln	Ala	Gln	Gly	Leu	Gly	Ser	
				245						250					255		
	Asn	Leu	Val	Thr	Glu	Val	Arg	Val	Tyr	Asn	Trp	Phe	Ala	Asn	Arg	Arg	
45				260					265					270			
	Lys	Glu	Glu	Ala	Phe	Arg	His	Lys	Leu	Ala	Met	Asp	Thr	Tyr	Ser	Gly	
		275						280					285				
50	Pro	Pro	Pro	Gly	Pro	Gly	Pro	Gly	Pro	Ala	Leu	Pro	Ala	His	Ser	Ser	
		290					295						300				
	Pro	Gly	Leu	Pro	Pro	Pro	Ala	Leu	Ser	Pro	Ser	Lys	Val	His	Gly	Val	
	305					310					315					320	
55	Arg	Gly	Gln	Pro	Ala	Thr	Ser	Glu	Thr	Ala	Glu	Val	Pro	Ser	Ser	Ser	
				325						330					335		
	Gly	Gly	Pro	Leu	Val	Thr	Val	Ser	Thr	Pro	Leu	His	Gln	Val	Ser	Pro	
60				340					345					350			
	Thr	Gly	Leu	Glu	Pro	Ser	His	Ser	Leu	Leu	Ser	Thr	Glu	Ala	Lys	Leu	

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3238 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
(B) LOCATION: 988
(D) OTHER INFORMATION: /mod_base= OTHER
/note= "N = A, C, G, or T"

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: join(24..986, 990..1916)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

```
CGTGGCCCTG TGGCAGCCGA GCC ATG GTT TCT AAA CTG AGC CAG CTG CAG      50
           Met Val Ser Lys Leu Ser Gln Leu Gln
           1                      5

ACG GAG CTC CTG GCG GCC CTG CTC GAG TCA GGG CTG AGC AAA GAG GCA      98
Thr Glu Leu Leu Ala Ala Leu Leu Glu Ser Gly Leu Ser Lys Glu Ala
 10                      15                      20                      25

CTG ATC CAG GCA CTG GGT GAG CCG GGG CCC TAC CTC CTG GCT GGA GAA     146
Leu Ile Gln Ala Leu Gly Glu Pro Gly Pro Tyr Leu Leu Ala Gly Glu
          30                      35                      40

GGC CCC CTG GAC AAG GGG GAG TCC TGC GGC GGC GGT CGA GGG GAG CTG     194
Gly Pro Leu Asp Lys Gly Glu Ser Cys Gly Gly Gly Arg Gly Glu Leu
          45                      50                      55

GCT GAG CTG CCC AAT GGG CTG GGG GAG ACT CGG GGC TCC GAG GAC GAG     242
Ala Glu Leu Pro Asn Gly Leu Gly Glu Thr Arg Gly Ser Glu Asp Glu
          60                      65                      70

ACG GAC GAC GAT GGG GAA GAC TTC ACG CCA CCC ATC CTC AAA GAG CTG     290
Thr Asp Asp Asp Gly Glu Asp Phe Thr Pro Pro Ile Leu Lys Glu Leu
          75                      80                      85

GAG AAC CTC AGC CCT GAG GAG GCG GCC CAC CAG AAA GCC GTG GTG GAG     338
Glu Asn Leu Ser Pro Glu Glu Ala Ala His Gln Lys Ala Val Val Glu
          90                      95                      100                      105

ACC CTT CTG CAG GAG GAC CCG TGG CGT GTG GCG AAG ATG GTC AAG TCC     386
Thr Leu Leu Gln Glu Asp Pro Trp Arg Val Ala Lys Met Val Lys Ser
          110                      115                      120

TAC CTG CAG CAG CAC AAC ATC CCA CAG CAG GAG GTG GTC GAT ACC ACT     434
Tyr Leu Gln Gln His Asn Ile Pro Gln Gln Glu Val Val Asp Thr Thr
          125                      130                      135

GGC CTC AAC CAG TCC CAC CTG TCC CAA CAC CTC AAC AAG GGC ACT CCC     482
Gly Leu Asn Gln Ser His Leu Ser Gln His Leu Asn Lys Gly Thr Pro
          140                      145                      150
```


5	ATG AAG ACG CAG AAG CGG GCC GCC CTG TAC ACC TGG TAC GTC CGC AAG Met Lys Thr Gln Lys Arg Ala Ala Leu Tyr Thr Trp Tyr Val Arg Lys 155 160 165	530
10	CAG CGA GAG GTG GCG CAG CAG TTC ACC CAT GCA GGG CAG GGA GGG CTG Gln Arg Glu Val Ala Gln Gln Phe Thr His Ala Gly Gln Gly Gly Leu 170 175 180 185	578
15	ATT GAA GAG CCC ACA GGT GAT GAG CTA CCA ACC AAG AAG GGG CGG AGG Ile Glu Glu Pro Thr Gly Asp Glu Leu Pro Thr Lys Lys Gly Arg Arg 190 195 200	626
20	AAC CGT TTC AAG TGG GGC CCA GCA TCC CAG CAG ATC CTG TTC CAG GCC Asn Arg Phe Lys Trp Gly Pro Ala Ser Gln Gln Ile Leu Phe Gln Ala 205 210 215	674
25	TAT GAG AGG CAG AAG AAC CCT AGC AAG GAG GAG CGA GAG ACG CTA GTG Tyr Glu Arg Gln Lys Asn Pro Ser Lys Glu Glu Arg Glu Thr Leu Val 220 225 230	722
30	GAG GAG TGC AAT AGG GCG GAA TGC ATC CAG AGA GGG GTG TCC CCA TCA Glu Glu Cys Asn Arg Ala Glu Cys Ile Gln Arg Gly Val Ser Pro Ser 235 240 245	770
35	CAG GCA CAG GGG CTG GGC TCC AAC CTC GTC ACG GAG GTG CGT GTC TAC Gln Ala Gln Gly Leu Gly Ser Asn Leu Val Thr Glu Val Arg Val Tyr 250 255 260 265	818
40	AAC TGG TTT GCC AAC CGG CGC AAA GAA GAA GCC TTC CGG CAC AAG CTG Asn Trp Phe Ala Asn Arg Arg Lys Glu Glu Ala Phe Arg His Lys Leu 270 275 280	866
45	GCC ATG GAC ACG TAC AGC GGG CCC CCC CCA GGG CCA GGC CCG GGA CCT Ala Met Asp Thr Tyr Ser Gly Pro Pro Pro Gly Pro Gly Pro Gly Pro 285 290 295	914
50	GCG CTG CCC GCT CAC AGC TCC CCT GGC CTG CCT CCA CCT GCC CTC TCC Ala Leu Pro Ala His Ser Ser Pro Gly Leu Pro Pro Pro Ala Leu Ser 300 305 310	962
55	CCC AGT AAG GTC CAC GGT GTG CGC TNT GGA CAG CCT GCG ACC AGT GAG Pro Ser Lys Val His Gly Val Arg Gly Gln Pro Ala Thr Ser Glu 315 320 325	1010
60	ACT GCA GAA GTA CCC TCA AGC AGC GGC GGT CCC TTA GTG ACA GTG TCT Thr Ala Glu Val Pro Ser Ser Ser Gly Gly Pro Leu Val Thr Val Ser 330 335 340	1058
65	ACA CCC CTC CAC CAA GTG TCC CCC ACG GGC CTG GAG CCC AGC CAC AGC Thr Pro Leu His Gln Val Ser Pro Thr Gly Leu Glu Pro Ser His Ser 345 350 355 360	1106
70	CTG CTG AGT ACA GAA GCC AAG CTG GTC TCA GCA GCT GGG GGC CCC CTC Leu Leu Ser Thr Glu Ala Lys Leu Val Ser Ala Ala Gly Gly Pro Leu 365 370 375	1154
75	CCC CCT GTC AGC ACC CTG ACA GCA CTG CAC AGC TTG GAG CAG ACA TCC Pro Pro Val Ser Thr Leu Thr Ala Leu His Ser Leu Glu Gln Thr Ser 380 385 390	1202

5
 10
 15
 20
 25
 30
 35
 40
 45

AGCCAGCCCT	GCCTGGAGGA	CCTGAGCCTG	CCGAGCAACC	GTGGCCCTTC	CTGGACAGCT	2036
GTGCCTCGCT	CCCCACTCTG	CTCTGATGCA	TCAGAAAGGG	AGGGCTCTGA	GGCGCCCCAA	2096
CCCGTGGAGG	CTGCTCGGGG	TGCACAGGAG	GGGGTCGTGG	AGAGCTAGGA	GCAAAGCCTG	2156
TTCATGGCAG	ATGTAGGAGG	GACTGTGCTG	GCTTCGTGGG	ATACAGTCTT	CTTACTTGGA	2216
ACTGAAGGGG	GCGGCCATATG	ACTTGGGCAC	CCCCAGCCTG	GGCCTATGGA	GAGCCCTGGG	2276
ACCGCTACAC	CACTCTGGCA	GCCACACTTC	TCAGGACACA	GGCCTGTGTA	GCTGTGACCT	2336
GCTGAGCTCT	GAGAGGCCCT	GGATCAGCGT	GGCCTTGTTT	TGTCACCAAT	GTACCCACCG	2396
GGCCACTCCT	TCCTGCCCCA	ACTCCTTCCA	GCTAGTGACC	CACATGCCAT	TTGTACTGAC	2456
CCCATCACCT	ACTCACACAG	GCATTTCTTG	GGTGGCTACT	CTGTGCCAGA	GCCTGGGGCT	2516
CTAACTGCCT	GAGCCCAGGG	AGGCCGAAGC	TAACAGGGAA	GGCAGGCAGG	GCTCTCCTGG	2576
TCTTCCCATC	CCCAGCGATT	CCCTCTCCCA	GGCCCCATGA	CCTCCAGCTT	TCCTGTATTT	2636
CTTCCCAAGA	GCATGATGCC	TCTGAGGCCA	GCCTGGCCTC	CTGCCTCTAC	TGGGAAGGCT	2696
ACTTCGGGGC	TGGGAAGTCG	TCCTTACTCC	TGTGGGAGCC	TCGCAACCCG	TGCCAAGTCC	2756
AGGTCCTGGT	GGGGCAGCTC	CTCTGTCTCG	AGCGCCCTGC	AGACCCTGCC	CTTGTTTGGG	2816
GCAGGAGTAG	CTGAGCTCAC	AAGGCAGCAA	GGCCCCGAGC	GCTGAGCAGG	GCCGGGGAAC	2876
TGGCCAAGCT	GAGGTGCCCA	GGAGAAGAAA	GAGGTGACCC	CAGGGCACAG	GAGCTACCTG	2936
TGTGGACAGG	ACTAACACTC	AGAAGCCTGG	GTGCCTGGCT	GGCTGAGGGC	AGTTCGCAGC	2996
CACCCTGAGG	AGTCTGAGGT	CCTGAGCACT	GCCAGGAGGG	ACAAAGGAGC	CTGTGAACCC	3056
AGGACAAGCA	TGGTCCCACA	TCCCTGGGCC	TGCTGCTGAG	AACCTGGCCT	TCAGTGTACC	3116
GCGTCTACCC	TGGGATTGAG	GAAAAGGCCT	GGGGTGACCC	GGCACCCCTT	GCAGCTTGTA	3176
GCCAGCCGGG	GCGAGTGGCA	CGTTTATTTA	ACTTTTAGTA	AAGTCAAGGA	GAAATGCGGT	3236
GG						3238

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 630 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met	Val	Ser	Lys	Leu	Ser	Gln	Leu	Gln	Thr	Glu	Leu	Leu	Ala	Ala	Leu
1				5				10					15		
Leu	Glu	Ser	Gly	Leu	Ser	Lys	Glu	Ala	Leu	Ile	Gln	Ala	Leu	Gly	Glu

20 25 30
 5 Pro Gly Pro Tyr Leu Leu Ala Gly Glu Gly Pro Leu Asp Lys Gly Glu
 35 40 45
 Ser Cys Gly Gly Gly Arg Gly Glu Leu Ala Glu Leu Pro Asn Gly Leu
 50 55 60
 10 Gly Glu Thr Arg Gly Ser Glu Asp Glu Thr Asp Asp Asp Gly Glu Asp
 65 70 75 80
 Phe Thr Pro Pro Ile Leu Lys Glu Leu Glu Asn Leu Ser Pro Glu Glu
 85 90 95
 15 Ala Ala His Gln Lys Ala Val Val Glu Thr Leu Leu Gln Glu Asp Pro
 100 105 110
 Trp Arg Val Ala Lys Met Val Lys Ser Tyr Leu Gln Gln His Asn Ile
 115 120 125
 20 Pro Gln Gln Glu Val Val Asp Thr Thr Gly Leu Asn Gln Ser His Leu
 130 135 140
 Ser Gln His Leu Asn Lys Gly Thr Pro Met Lys Thr Gln Lys Arg Ala
 145 150 155 160
 Ala Leu Tyr Thr Trp Tyr Val Arg Lys Gln Arg Glu Val Ala Gln Gln
 165 170 175
 30 Phe Thr His Ala Gly Gln Gly Gly Leu Ile Glu Glu Pro Thr Gly Asp
 180 185 190
 Glu Leu Pro Thr Lys Lys Gly Arg Arg Asn Arg Phe Lys Trp Gly Pro
 195 200 205
 35 Ala Ser Gln Gln Ile Leu Phe Gln Ala Tyr Glu Arg Gln Lys Asn Pro
 210 215 220
 Ser Lys Glu Glu Arg Glu Thr Leu Val Glu Glu Cys Asn Arg Ala Glu
 225 230 235 240
 40 Cys Ile Gln Arg Gly Val Ser Pro Ser Gln Ala Gln Gly Leu Gly Ser
 245 250 255
 45 Asn Leu Val Thr Glu Val Arg Val Tyr Asn Trp Phe Ala Asn Arg Arg
 260 265 270
 Lys Glu Glu Ala Phe Arg His Lys Leu Ala Met Asp Thr Tyr Ser Gly
 275 280 285
 50 Pro Pro Pro Gly Pro Gly Pro Gly Pro Ala Leu Pro Ala His Ser Ser
 290 295 300
 55 Pro Gly Leu Pro Pro Pro Ala Leu Ser Pro Ser Lys Val His Gly Val
 305 310 315 320
 Arg Gly Gln Pro Ala Thr Ser Glu Thr Ala Glu Val Pro Ser Ser Ser
 325 330 335
 60 Gly Gly Pro Leu Val Thr Val Ser Thr Pro Leu His Gln Val Ser Pro
 340 345 350

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3239 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
(B) LOCATION: 989
(D) OTHER INFORMATION: /mod_base= OTHER

/note= "N = A, C, G, or T"

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 24..965

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```
CGTGGCCCTG TGGCAGCCGA GCC ATG GTT TCT AAA CTG AGC CAG CTG CAG      50
      Met Val Ser Lys Leu Ser Gln Leu Gln
      1              5

ACG GAG CTC CTG GCG GCC CTG CTC GAG TCA GGG CTG AGC AAA GAG GCA      98
Thr Glu Leu Leu Ala Ala Leu Leu Glu Ser Gly Leu Ser Lys Glu Ala
 10              15              20              25

CTG ATC CAG GCA CTG GGT GAG CCG GGG CCC TAC CTC CTG GCT GGA GAA     146
Leu Ile Gln Ala Leu Gly Glu Pro Gly Pro Tyr Leu Leu Ala Gly Glu
      30              35              40

GGC CCC CTG GAC AAG GGG GAG TCC TGC GGC GGC GGT CGA GGG GAG CTG     194
Gly Pro Leu Asp Lys Gly Glu Ser Cys Gly Gly Gly Arg Gly Glu Leu
      45              50              55

GCT GAG CTG CCC AAT GGG CTG GGG GAG ACT CGG GGC TCC GAG GAC GAG     242
Ala Glu Leu Pro Asn Gly Leu Gly Glu Thr Arg Gly Ser Glu Asp Glu
      60              65              70

ACG GAC GAC GAT GGG GAA GAC TTC ACG CCA CCC ATC CTC AAA GAG CTG     290
Thr Asp Asp Asp Gly Glu Asp Phe Thr Pro Pro Ile Leu Lys Glu Leu
      75              80              85

GAG AAC CTC AGC CCT GAG GAG GCG GCC CAC CAG AAA GCC GTG GTG GAG     338
Glu Asn Leu Ser Pro Glu Glu Ala Ala His Gln Lys Ala Val Val Glu
      90              95              100              105

ACC CTT CTG CAG GAG GAC CCG TGG CGT GTG GCG AAG ATG GTC AAG TCC     386
Thr Leu Leu Gln Glu Asp Pro Trp Arg Val Ala Lys Met Val Lys Ser
      110              115              120

TAC CTG CAG CAG CAC AAC ATC CCA CAG CGG GAG GTG GTC GAT ACC ACT     434
Tyr Leu Gln Gln His Asn Ile Pro Gln Arg Glu Val Val Asp Thr Thr
      125              130              135

GGC CTC AAC CAG TCC CAC CTG TCC CAA CAC CTC AAC AAG GGC ACT CCC     482
Gly Leu Asn Gln Ser His Leu Ser Gln His Leu Asn Lys Gly Thr Pro
      140              145              150
```

5	ATG AAG ACG CAG AAG CGG GCC GCC CTG TAC ACC TGG TAC GTC CGC AAG Met Lys Thr Gln Lys Arg Ala Ala Leu Tyr Thr Trp Tyr Val Arg Lys 155 160 165	530
10	CAG CGA GAG GTG GCG CAG CAG TTC ACC CAT GCA GGG CAG GGA GGG CTG Gln Arg Glu Val Ala Gln Gln Phe Thr His Ala Gly Gln Gly Gly Leu 170 175 180 185	578
15	ATT GAA GAG CCC ACA GGT GAT GAG CTA CCA ACC AAG AAG GGG CGG AGG Ile Glu Glu Pro Thr Gly Asp Glu Leu Pro Thr Lys Lys Gly Arg Arg 190 195 200	626
20	AAC CGT TTC AAG TGG GGC CCA GCA TCC CAG CAG ATC CTG TTC CAG GCC Asn Arg Phe Lys Trp Gly Pro Ala Ser Gln Gln Ile Leu Phe Gln Ala 205 210 215	674
25	TAT GAG AGG CAG AAG AAC CCT AGC AAG GAG GAG CGA GAG ACG CTA GTG Tyr Glu Arg Gln Lys Asn Pro Ser Lys Glu Glu Arg Glu Thr Leu Val 220 225 230	722
30	GAG GAG TGC AAT AGG GCG GAA TGC ATC CAG AGA GGG GTG TCC CCA TCA Glu Glu Cys Asn Arg Ala Glu Cys Ile Gln Arg Gly Val Ser Pro Ser 235 240 245	770
35	CAG GCA CAG GGG CTG GGC TCC AAC CTC GTC ACG GAG GTG CGT GTC TAC Gln Ala Gln Gly Leu Gly Ser Asn Leu Val Thr Glu Val Arg Val Tyr 250 255 260 265	818
40	AAC TGG TTT GCC AAC CGG CGC AAA GAA GAA GCC TTC CGG CAC AAG CTG Asn Trp Phe Ala Asn Arg Arg Lys Glu Glu Ala Phe Arg His Lys Leu 270 275 280	866
45	GCC ATG GAC ACG TAC AGC GGG CCC CCC CCC AGG GCC AGG CCC GGG ACC Ala Met Asp Thr Tyr Ser Gly Pro Pro Pro Arg Ala Arg Pro Gly Thr 285 290 295	914
50	TGC GCT GCC CGC TCA CAG CTC CCC TGG CCT GCC TCC ACC TGC CCT CTC Cys Ala Ala Arg Ser Gln Leu Pro Trp Pro Ala Ser Thr Cys Pro Leu 300 305 310	962
55	CCC CAGTAAGGTC CACGGTGTGC GCTNTGGACA GCCTGCGACC AGTGAGACTG Pro	1015
60	CAGAAGTACC CTCAAGCAGC GGCGGTCCCT TAGTGACAGT GTCTACACCC CTCCACCAAG TGTCCTCCAC GGGCCTGGAG CCCAGCCACA GCCTGCTGAG TACAGAAGCC AAGCTGGTCT CAGCAGCTGG GGGCCCCCTC CCCCCTGTCA GCACCTGAC AGCACTGCAC AGCTTGAGC AGACATCCCC AGGCCTCAAC CAGCAGCCCC AGAACCTCAT CATGGCCTCA CTTCTGGGG TCATGACCAT CGGGCCTGGT GAGCCTGCCT CCCTGGGTCC TACGTTACAC AACACAGGTG CCTCCACCCT GGTCATCGGC CTGGCCTCCA CGCAGGCACA GAGTGTGCCG GTCATCAACA GCATGGGCAG CAGCCTGACC ACCCTGCAGC CCGTCCAGTT CTCCCAGCCG CTGCACCCCT CCTACCAGCA GCCGCTCATG CCACCTGTGC AGAGCCATGT GACCCAGAGC CCCTTCATGG	1075 1135 1195 1255 1315 1375 1435 1495

	CCACCATGGC TCAGCTGCAG AGCCCCACG CCCTCTACAG CCACAAGCCC GAGGTGGCCC	1555
	AGTACACCCA CACGGGCTTG CTCCCGCAGA CTATGCTCAT CACCGACACC ACCAACCTGA	1615
5	GCGCCCTGGC CAGCCTCACG CCCACCAAGC AGGTCTTCAC CTCAGACACT GAGGCCTCCA	1675
	GTGAGTCCGG GCTTCACACG CCGGCATCTC AGGCCACCAC CCTCCACGTC CCCAGCCAGG	1735
	ACCCTGCCGG CATCCAGCAC CTGCAGCCGG CCCACCGGCT CAGCGCCAGC CCCACAGTGT	1795
10	CCTCCAGCAG CCTGGTGCTG TACCAGAGCT CAGACTCCAG CAATGGCCAG AGCCACCTGC	1855
	TGCCATCCAA CCACAGCGTC ATCGAGACCT TCATCTCCAC CCAGATGGCC TCTTCCTCCC	1915
15	AGTAACCACG GCACCTGGGC CCTGGGGCCT GTACTGCCTG CTTGGGGGGT GATGAGGGCA	1975
	GCAGCCAGCC CTGCCTGGAG GACCTGAGCC TGCCGAGCAA CCGTGGCCCT TCCTGGACAG	2035
	CTGTGCCTCG CTCCCCACTC TGCTCTGATG CATCAGAAAG GGAGGGCTCT GAGGCGCCCC	2095
20	AACCCGTGGA GGCTGCTCGG GGTGCACAGG AGGGGGTCGT GGAGAGCTAG GAGCAAAGCC	2155
	TGTTTCATGGC AGATGTAGGA GGGACTGTCG CTGCTTCGTG GGATACAGTC TTCTTACTTG	2215
25	GAAGTGAAGG GGGCGGCCTA TGACTTGGGC ACCCCCAGCC TGGGCCTATG GAGAGCCCTG	2275
	GGACCGCTAC ACCACTCTGG CAGCCACACT TCTCAGGACA CAGGCCTGTG TAGCTGTGAC	2335
	CTGCTGAGCT CTGAGAGGCC CTGGATCAGC GTGGCCTTGT TCTGTCACCA ATGTACCCAC	2395
30	CGGGCCACTC CTTCTGCCCC CAACTCCTTC CAGCTAGTGA CCCACATGCC ATTTGTAAGT	2455
	ACCCCATCAC CTACTCACAC AGGCATTTCC TGGGTGGCTA CTCTGTGCCA GAGCCTGGGG	2515
35	CTCTAACTGC CTGAGCCCAG GGAGGCCGAA GCTAACAGGG AAGGCAGGCA GGGCTCTCCT	2575
	GGTCTTCCCA TCCCCAGCGA TTCCCTCTCC CAGGCCCCAT GACCTCCAGC TTTCTGTAT	2635
	TTCTTCCCAA GAGCATGATG CCTCTGAGGC CAGCCTGGCC TCCTGCCTCT ACTGGAAGG	2695
40	CTACTTCGGG GCTGGGAAGT CGTCCTTACT CCTGTGGGAG CCTCGCAACC CGTGCCAAGT	2755
	CCAGGTCCTG GTGGGGCAGC TCCTCTGTCT CGAGCGCCCT GCAGACCCTG CCCTGTGTTG	2815
45	GGGCAGGAGT AGCTGAGCTC ACAAGGCAGC AAGGCCCGAG CAGCTGAGCA GGGCCGGGGA	2875
	ACTGGCCAAG CTGAGGTGCC CAGGAGAAGA AAGAGGTGAC CCCAGGGCAC AGGAGCTACC	2935
	TGTGTGGACA GGAATAACAC TCAGAAGCCT GGGTGCCTGG CTGGCTGAGG GCAGTTCGCA	2995
50	GCCACCCTGA GGAGTCTGAG GTCCTGAGCA CTGCCAGGAG GGACAAAGGA GCCTGTGAAC	3055
	CCAGGACAAG CATGGTCCCA CATCCCTGGG CCTGCTGCTG AGAACCTGGC CTTCAAGTGA	3115
55	CCGCGTCTAC CCTGGGATTC AGGAAAAGGC CTGGGGTGAC CCGGCACCCC CTGCAGCTTG	3175
	TAGCCAGCCG GGGCGAGTGG CACGTTTATT TAACTTTTAG TAAAGTCAAG GAGAAATGCG	3235
60	GTGA	3239

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 314 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Val Ser Lys Leu Ser Gln Leu Gln Thr Glu Leu Leu Ala Ala Leu
1 5 10 15
Leu Glu Ser Gly Leu Ser Lys Glu Ala Leu Ile Gln Ala Leu Gly Glu
20 25 30
Pro Gly Pro Tyr Leu Leu Ala Gly Glu Gly Pro Leu Asp Lys Gly Glu
35 40 45
Ser Cys Gly Gly Gly Arg Gly Glu Leu Ala Glu Leu Pro Asn Gly Leu
50 55 60
Gly Glu Thr Arg Gly Ser Glu Asp Glu Thr Asp Asp Asp Gly Glu Asp
65 70 75 80
Phe Thr Pro Pro Ile Leu Lys Glu Leu Glu Asn Leu Ser Pro Glu Glu
85 90 95
Ala Ala His Gln Lys Ala Val Val Glu Thr Leu Leu Gln Glu Asp Pro
100 105 110
Trp Arg Val Ala Lys Met Val Lys Ser Tyr Leu Gln Gln His Asn Ile
115 120 125
Pro Gln Arg Glu Val Val Asp Thr Thr Gly Leu Asn Gln Ser His Leu
130 135 140
Ser Gln His Leu Asn Lys Gly Thr Pro Met Lys Thr Gln Lys Arg Ala
145 150 155 160
Ala Leu Tyr Thr Trp Tyr Val Arg Lys Gln Arg Glu Val Ala Gln Gln
165 170 175
Phe Thr His Ala Gly Gln Gly Gly Leu Ile Glu Glu Pro Thr Gly Asp
180 185 190
Glu Leu Pro Thr Lys Lys Gly Arg Arg Asn Arg Phe Lys Trp Gly Pro
195 200 205
Ala Ser Gln Gln Ile Leu Phe Gln Ala Tyr Glu Arg Gln Lys Asn Pro
210 215 220
Ser Lys Glu Glu Arg Glu Thr Leu Val Glu Glu Cys Asn Arg Ala Glu
225 230 235 240
Cys Ile Gln Arg Gly Val Ser Pro Ser Gln Ala Gln Gly Leu Gly Ser
245 250 255
Asn Leu Val Thr Glu Val Arg Val Tyr Asn Trp Phe Ala Asn Arg Arg
260 265 270

Lys Glu Glu Ala Phe Arg His Lys Leu Ala Met Asp Thr Tyr Ser Gly
 275 280 285

Pro Pro Pro Arg Ala Arg Pro Gly Thr Cys Ala Ala Arg Ser Gln Leu
 290 295 300

Pro Trp Pro Ala Ser Thr Cys Pro Leu Pro
 305 310

(2) INFORMATION FOR SEQ ID NO:7:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 3236 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
 - (B) LOCATION: 988
 - (D) OTHER INFORMATION: /mod_base= OTHER
- /note= "N = A, C, G, or T"

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(24..986, 990..1271)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

CGTGGCCCTG TGGCAGCCGA GCC ATG GTT TCT AAA CTG AGC CAG CTG CAG	50
Met Val Ser Lys Leu Ser Gln Leu Gln	
1 5	
ACG GAG CTC CTG GCG GCC CTG CTC GAG TCA GGG CTG AGC AAA GAG GCA	98
Thr Glu Leu Leu Ala Ala Leu Leu Glu Ser Gly Leu Ser Lys Glu Ala	
10 15 20 25	
CTG ATC CAG GCA CTG GGT GAG CCG GGG CCC TAC CTC CTG GCT GGA GAA	146
Leu Ile Gln Ala Leu Gly Glu Pro Gly Pro Tyr Leu Leu Ala Gly Glu	
30 35 40	
GGC CCC CTG GAC AAG GGG GAG TCC TGC GGC GGC GGT CGA GGG GAG CTG	194
Gly Pro Leu Asp Lys Gly Glu Ser Cys Gly Gly Gly Arg Gly Glu Leu	
45 50 55	
GCT GAG CTG CCC AAT GGG CTG GGG GAG ACT CGG GGC TCC GAG GAC GAG	242
Ala Glu Leu Pro Asn Gly Leu Gly Glu Thr Arg Gly Ser Glu Asp Glu	
60 65 70	
ACG GAC GAC GAT GGG GAA GAC TTC ACG CCA CCC ATC CTC AAA GAG CTG	290
Thr Asp Asp Asp Gly Glu Asp Phe Thr Pro Pro Ile Leu Lys Glu Leu	
75 80 85	
GAG AAC CTC AGC CCT GAG GAG GCG GCC CAC CAG AAA GCC GTG GTG GAG	338
Glu Asn Leu Ser Pro Glu Glu Ala Ala His Gln Lys Ala Val Val Glu	
90 95 100 105	

	ACC	CTT	CTG	CAG	GAG	GAC	CCG	TGG	CGT	GTG	GCG	AAG	ATG	GTC	AAG	TCC	386
	Thr	Leu	Leu	Gln	Glu	Asp	Pro	Trp	Arg	Val	Ala	Lys	Met	Val	Lys	Ser	
					110					115					120		
5	TAC	CTG	CAG	CAG	CAC	AAC	ATC	CCA	CAG	CGG	GAG	GTG	GTC	GAT	ACC	ACT	434
	Tyr	Leu	Gln	Gln	His	Asn	Ile	Pro	Gln	Arg	Glu	Val	Val	Asp	Thr	Thr	
				125					130					135			
10	GGC	CTC	AAC	CAG	TCC	CAC	CTG	TCC	CAA	CAC	CTC	AAC	AAG	GGC	ACT	CCC	482
	Gly	Leu	Asn	Gln	Ser	His	Leu	Ser	Gln	His	Leu	Asn	Lys	Gly	Thr	Pro	
			140					145					150				
15	ATG	AAG	ACG	CAG	AAG	CGG	GCC	GCC	CTG	TAC	ACC	TGG	TAC	GTC	CGC	AAG	530
	Met	Lys	Thr	Gln	Lys	Arg	Ala	Ala	Leu	Tyr	Thr	Trp	Tyr	Val	Arg	Lys	
		155					160					165					
20	CAG	CGA	GAG	GTG	GCG	CAG	CAG	TTC	ACC	CAT	GCA	GGG	CAG	GGA	GGG	CTG	578
	Gln	Arg	Glu	Val	Ala	Gln	Gln	Phe	Thr	His	Ala	Gly	Gln	Gly	Gly	Leu	
	170					175					180					185	
25	ATT	GAA	GAG	CCC	ACA	GGT	GAT	GAG	CTA	CCA	ACC	AAG	AAG	GGG	CGG	AGG	626
	Ile	Glu	Glu	Pro	Thr	Gly	Asp	Glu	Leu	Pro	Thr	Lys	Lys	Gly	Arg	Arg	
					190					195					200		
30	AAC	CGT	TTC	AAG	TGG	GGC	CCA	GCA	TCC	CAG	CAG	ATC	CTG	TTC	CAG	GCC	674
	Asn	Arg	Phe	Lys	Trp	Gly	Pro	Ala	Ser	Gln	Gln	Ile	Leu	Phe	Gln	Ala	
				205					210					215			
35	TAT	GAG	AGG	CAG	AAG	AAC	CCT	AGC	AAG	GAG	GAG	CGA	GAG	ACG	CTA	GTG	722
	Tyr	Glu	Arg	Gln	Lys	Asn	Pro	Ser	Lys	Glu	Glu	Arg	Glu	Thr	Leu	Val	
			220					225					230				
40	GAG	GAG	TGC	AAT	AGG	GCG	GAA	TGC	ATC	CAG	AGA	GGG	GTG	TCC	CCA	TCA	770
	Glu	Glu	Cys	Asn	Arg	Ala	Glu	Cys	Ile	Gln	Arg	Gly	Val	Ser	Pro	Ser	
		235					240					245					
45	CAG	GCA	CAG	GGG	CTG	GGC	TCC	AAC	CTC	GTC	ACG	GAG	GTG	CGT	GTC	TAC	818
	Gln	Ala	Gln	Gly	Leu	Gly	Ser	Asn	Leu	Val	Thr	Glu	Val	Arg	Val	Tyr	
	250					255					260					265	
50	AAC	TGG	TTT	GCC	AAC	CGG	CGC	AAA	GAA	GAA	GCC	TTC	CGG	CAC	AAG	CTG	866
	Asn	Trp	Phe	Ala	Asn	Arg	Arg	Lys	Glu	Glu	Ala	Phe	Arg	His	Lys	Leu	
				270					275						280		
55	GCC	ATG	GAC	ACG	TAC	AGC	GGG	CCC	CCC	CCA	GGG	CCA	GGC	CCG	GGA	CCT	914
	Ala	Met	Asp	Thr	Tyr	Ser	Gly	Pro	Pro	Pro	Gly	Pro	Gly	Pro	Gly	Pro	
				285					290					295			
60	GCG	CTG	CCC	GCT	CAC	AGC	TCC	CCT	GGC	CTG	CCT	CCA	CCT	GCC	CTC	TCC	962
	Ala	Leu	Pro	Ala	His	Ser	Ser										

	Thr	Pro	Leu	His	Gln	Val	Ser	Pro	Thr	Gly	Leu	Glu	Pro	Ser	His	Ser	
	345					350					355					360	
5	CTG	CTG	AGT	ACA	GAA	GCC	AAG	CTG	GTC	TCA	GCA	GCT	GGG	GGC	CCC	CTC	1154
	Leu	Leu	Ser	Thr	Glu	Ala	Lys	Leu	Val	Ser	Ala	Ala	Gly	Gly	Pro	Leu	
					365					370					375		
10	CCC	CGT	CAG	CAC	CCT	GAC	AGC	ACT	GCA	CAG	CTT	GGA	GCA	GAC	ATC	CCC	1202
	Pro	Arg	Gln	His	Pro	Asp	Ser	Thr	Ala	Gln	Leu	Gly	Ala	Asp	Ile	Pro	
				380					385					390			
15	AGG	CCT	CAA	CCA	GCA	GCC	CCA	GAA	CCT	CAT	CAT	GGC	CTC	ACT	TCC	TGG	1250
	Arg	Pro	Gln	Pro	Ala	Ala	Pro	Glu	Pro	His	His	Gly	Leu	Thr	Ser	Trp	
			395					400					405				
	GGT	CAT	GAC	CAT	CGG	GCC	TGG	TGAGCCTGCC	TCCCTGGGTC	CTACGTTTAC							1301
	Gly	His	Asp	His	Arg	Ala	Trp										
	410					415											
20	CAACACAGGT	GCCTCCACCC	TGGTCATCGG	CCTGGCCTCC	ACGCAGGCAC	AGAGTGTGCC											1361
	GGTCATCAAC	AGCATGGGCA	GCAGCCTGAC	CACCCTGCAG	CCCGTCCAGT	TCTCCCAGCC											1421
25	GCTGCACCCC	TCCTACCAGC	AGCCGCTCAT	GCCACCTGTG	CAGAGCCATG	TGACCCAGAG											1481
	CCCCTTCATG	GCCACCATGG	CTCAGCTGCA	GAGCCCCAC	GCCCTCTACA	GCCACAAGCC											1541
	CGAGGTGGCC	CAGTACACCC	ACACGGGCCT	GCTCCCGCAG	ACTATGCTCA	TCACCGACAC											1601
30	CACCAACCTG	AGCGCCCTGG	CCAGCCTCAC	GCCCACCAAG	CAGGTCTTCA	CCTCAGACAC											1661
	TGAGGCCTCC	AGTGAGTCCG	GGCTTCACAC	GCCGGCATCT	CAGGCCACCA	CCCTCCACGT											1721
	CCCCAGCCAG	GACCCTGCCG	GCATCCAGCA	CCTGCAGCCG	GCCCACCGGC	TCAGCGCCAG											1781
35	CCCCACAGTG	TCCTCCAGCA	GCCTGGTGCT	GTACCAGAGC	TCAGACTCCA	GCAATGGCCA											1841
	GAGCCACCTG	CTGCCATCCA	ACCACAGCGT	CATCGAGACC	TTCATCTCCA	CCCAGATGGC											1901
40	CTCTTCCTCC	CAGTAACCAC	GGCACCTGGG	CCCTGGGGCC	TGTACTGCCT	GCTTGGGGGG											1961
	TGATGAGGGC	AGCAGCCAGC	CCTGCCTGGA	GGACCTGAGC	CTGCCGAGCA	ACCGTGGCCC											2021
	TTCTTGACA	GCTGTGCCTC	GCTCCCCACT	CTGCTCTGAT	GCATCAGAAA	GGGAGGGCTC											2081
45	TGAGGCGCCC	CAACCCGTGG	AGGCTGCTCG	GGGTGCACAG	GAGGGGGTCG	TGGAGAGCTA											2141
	GGAGCAAAGC	CTGTTCATGG	CAGATGTAGG	AGGGACTGTC	GCTGCTTCGT	GGGATACAGT											2201
50	CTTCTTACTT	GGAAGTGAAG	GGGGCGGCCT	ATGACTTGGG	CACCCCCAGC	CTGGGCCTAT											2261
	GGAGAGCCCT	GGGACCGCTA	CACCACTCTG	GCAGCCACAC	TTCTCAGGAC	ACAGGCCTGT											2321
	GTAGCTGTGA	CCTGCTGAGC	TCTGAGAGGC	CCTGGATCAG	CGTGGCCTTG	TTCTGTCACC											2381
55	AATGTACCCA	CCGGGCCACT	CCTTCCTGCC	CCAACTCCTT	CCAGCTAGTG	ACCCACATGC											2441
	CATTTGTACT	GACCCCATCA	CCTACTCACA	CAGGCATTTT	CTGGGTGGCT	ACTCTGTGCC											2501
60	AGAGCCTGGG	GCTCTAACTG	CCTGAGCCCA	GGGAGGCCGA	AGCTAACAGG	GAAGGCAGGC											2561

AGGGCTCTCC TGGTCTTCCC ATCCCCAGCG ATTCCCTCTC CCAGGCCCCA TGACCTCCAG 2621
 CTTTCCTGTA TTTCTTCCCA AGAGCATGAT GCCTCTGAGG CCAGCCTGGC CTCCTGCCTC 2681
 5 TACTGGGAAG GCTACTTCGG GGCTGGGAAG TCGTCCTTAC TCCTGTGGGA GCCTCGCAAC 2741
 CCGTGCCAAG TCCAGGTCCT GGTGGGGCAG CTCCTCTGTC TCGAGCGCCC TGCAGACCCT 2801
 10 GCCCTTGTTT GGGGCAGGAG TAGCTGAGCT CACAAGGCAG CAAGGCCCGA GCAGCTGAGC 2861
 AGGGCCGGGG AACTGGCCAA GCTGAGGTGC CCAGGAGAAG AAAGAGGTGA CCCCAGGGCA 2921
 CAGGAGCTAC CTGTGTGGAC AGGACTAACA CTCAGAAGCC TGGGTGCCTG GCTGGCTGAG 2981
 15 GGCAGTTCGC AGCCACCCTG AGGAGTCTGA GGTCTTGAGC ACTGCCAGGA GGGACAAAGG 3041
 AGCCTGTGAA CCCAGGACAA GCATGGTCCC ACATCCCTGG GCCTGCTGCT GAGAACCTGG 3101
 20 CCTTCAGTGT ACCGCGTCTA CCCTGGGATT CAGGAAAAGG CCTGGGGTGA CCCGGCACCC 3161
 CCTGCAGCTT GTAGCCAGCC GGGGCGAGTG GCACGTTTAT TTAACTTTGA GTAAAGTCAA 3221
 GGAGAAATGC GGTGG 3236

(2) INFORMATION FOR SEQ ID NO:8:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 415 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

Met Val Ser Lys Leu Ser Gln Leu Gln Thr Glu Leu Leu Ala Ala Leu
 1 5 10 15
 40 Leu Glu Ser Gly Leu Ser Lys Glu Ala Leu Ile Gln Ala Leu Gly Glu
 20 25 30
 Pro Gly Pro Tyr Leu Leu Ala Gly Glu Gly Pro Leu Asp Lys Gly Glu
 35 40 45
 45 Ser Cys Gly Gly Gly Arg Gly Glu Leu Ala Glu Leu Pro Asn Gly Leu
 50 55 60
 50 Gly Glu Thr Arg Gly Ser Glu Asp Glu Thr Asp Asp Asp Gly Glu Asp
 65 70 75 80
 Phe Thr Pro Pro Ile Leu Lys Glu Leu Glu Asn Leu Ser Pro Glu Glu
 85 90 95
 55 Ala Ala His Gln Lys Ala Val Val Glu Thr Leu Leu Gln Glu Asp Pro
 100 105 110
 Trp Arg Val Ala Lys Met Val Lys Ser Tyr Leu Gln Gln His Asn Ile
 115 120 125
 60 Pro Gln Arg Glu Val Val Asp Thr Thr Gly Leu Asn Gln Ser His Leu

	130		135		140														
5	Ser	Gln	His	Leu	Asn	Lys	Gly	Thr	Pro	Met	Lys	Thr	Gln	Lys	Arg	Ala			
	145					150					155					160			
	Ala	Leu	Tyr	Thr	Trp	Tyr	Val	Arg	Lys	Gln	Arg	Glu	Val	Ala	Gln	Gln			
					165					170					175				
10	Phe	Thr	His	Ala	Gly	Gln	Gly	Gly	Leu	Ile	Glu	Glu	Pro	Thr	Gly	Asp			
				180					185					190					
	Glu	Leu	Pro	Thr	Lys	Lys	Gly	Arg	Arg	Asn	Arg	Phe	Lys	Trp	Gly	Pro			
			195					200					205						
15	Ala	Ser	Gln	Gln	Ile	Leu	Phe	Gln	Ala	Tyr	Glu	Arg	Gln	Lys	Asn	Pro			
	210						215					220							
	Ser	Lys	Glu	Glu	Arg	Glu	Thr	Leu	Val	Glu	Glu	Cys	Asn	Arg	Ala	Glu			
20	225					230					235					240			
	Cys	Ile	Gln	Arg	Gly	Val	Ser	Pro	Ser	Gln	Ala	Gln	Gly	Leu	Gly	Ser			
					245					250					255				
25	Asn	Leu	Val	Thr	Glu	Val	Arg	Val	Tyr	Asn	Trp	Phe	Ala	Asn	Arg	Arg			
				260					265					270					
	Lys	Glu	Glu	Ala	Phe	Arg	His	Lys	Leu	Ala	Met	Asp	Thr	Tyr	Ser	Gly			
			275				280					285							
30	Pro	Pro	Pro	Gly	Pro	Gly	Pro	Gly	Pro	Ala	Leu	Pro	Ala	His	Ser	Ser			
	290					295						300							
	Pro	Gly	Leu	Pro	Pro	Pro	Ala	Leu	Ser	Pro	Ser	Lys	Val	His	Gly	Val			
35	305					310					315					320			
	Arg	Gly	Gln	Pro	Ala	Thr	Ser	Glu	Thr	Ala	Glu	Val	Pro	Ser	Ser	Ser			
					325					330					335				
40	Gly	Gly	Pro	Leu	Val	Thr	Val	Ser	Thr	Pro	Leu	His	Gln	Val	Ser	Pro			
				340					345					350					
	Thr	Gly	Leu	Glu	Pro	Ser	His	Ser	Leu	Leu	Ser	Thr	Glu	Ala	Lys	Leu			
			355					360					365						
45	Val	Ser	Ala	Ala	Gly	Gly	Pro	Leu	Pro	Arg	Gln	His	Pro	Asp	Ser	Thr			
	370						375					380							
	Ala	Gln	Leu	Gly	Ala	Asp	Ile	Pro	Arg	Pro	Gln	Pro	Ala	Ala	Pro	Glu			
	385					390					395					400			
50	Pro	His	His	Gly	Leu	Thr	Ser	Trp	Gly	His	Asp	His	Arg	Ala	Trp				
					405					410					415				

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 13 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: modified_base
 - (B) LOCATION: 7
 - (D) OTHER INFORMATION: /mod_base= OTHER
- /note= "N = A, C, G, or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

GTAAATNATT ACC

13

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

TACACCACTC TGGCAGCCAC ACT

23

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 24 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

CGGTGGGTAC ATTGGTGACA GAAC

24

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGCAGGCAAA CGCAACCCAC G

21

(2) INFORMATION FOR SEQ ID NO:13:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

GAAGGGGGGC TCGTTAGGAG C

21

(2) INFORMATION FOR SEQ ID NO:14:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

CATGCACAGT CCCCACCCTC A

21

(2) INFORMATION FOR SEQ ID NO:15:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTTCCAGCCC CCACCTATGA G

21

(2) INFORMATION FOR SEQ ID NO:16:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

GGGCAAGGTC AGGGGAATGG A

21

(2) INFORMATION FOR SEQ ID NO:17:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

CAGCCCAGAC CAAACCAGCA C

21

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

CAGAACCCTC CCCTTCATGC C

21

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGTGACTGCT GTCAATGGGA C

21

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

GGCAGACAGG CAGATGGCCT A

21

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

GCCTCCCTAG GGACTGCTCC A

21

(2) INFORMATION FOR SEQ ID NO:22:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

TGGAGCAGTC CCTAGGGAGG C

21

(2) INFORMATION FOR SEQ ID NO:23:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

GTGCCCCAT GAGCCTCCCA C

21

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

GGTCTTGGGC AGGGGTGGGA T

21

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

CTGCAATGCC TGCCAGGCAC C

21

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCCCTGCATC CATTGACAGC C

21

(2) INFORMATION FOR SEQ ID NO:27:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

GAGGCCTGGG ACTAGGGCTG T

21

(2) INFORMATION FOR SEQ ID NO:28:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CTCTGTCACA GGCCGAGGGA G

21

(2) INFORMATION FOR SEQ ID NO:29:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CCTGTGACAG AGCCCCTCAC C

21

(2) INFORMATION FOR SEQ ID NO:30:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

CGGACAGCAA CAGAAGGGGT G

21

(2) INFORMATION FOR SEQ ID NO:31:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

CAGAGCCCCT CACCCCACA T

21

(2) INFORMATION FOR SEQ ID NO:32:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

GTACCCCTAG GGACAGGCAG G

21

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

ACCCCCCAAG CAGGCAGTAC A

21

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 671 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 104..217

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

GCAGAGAGGG CACTGGGAGG AGGCAGTGGG AGGGCGGAGG GCGGGGGCCT TCGGGGTGGG
CGCCCAGGGT AGGGCAGGTG GCCGCGGCGT GGAGGCAGGG AGA ATG CGA CTC TCC

60

115

Met Arg Leu Ser
1

5 AAA ACC CTC GTC GAC ATG GAC ATG GCC GAC TAC AGT GCT GCA CTG GAC 163
Lys Thr Leu Val Asp Met Asp Met Ala Asp Tyr Ser Ala Ala Leu Asp
5 10 15 20

10 CCA GCC TAC ACC ACC CTG GAA TTT GAG AAT GTG CAG GTG TTG ACG ATG 211
Pro Ala Tyr Thr Thr Leu Glu Phe Glu Asn Val Gln Val Leu Thr Met
25 30 35

15 GGC AAT GGTAGGTGGG GGCAGATGTG CCCAGGTGTG CCAGTGGGGG CAGGTGTGCC 267
Gly Asn

20 TGGGTCCAGG AGCAGATCTT TGGCACTCAA CTTTGGGGTG GGAGGAGAAT GATACAAAAT 327

25 GGTAGGTTGG TCCTACAGGC CAGCACAGGT GTTGCCAAGT GAAGCCCATG TGCCCAGGCA 387

30 CAGTGATCAC AGGCATTCTG GGTGAAGGGA GGCCTGCAAG GGCCAATTTT CAGCAAAAGT 447

35 CGATCCCGGC TATTCCTCCC AGGCCCTTCC AGTCCTCACT GCCTCACAGT GGCTCTGCTT 507

40 GGCGCTTGGC ACAGTGACAT GATGGTGAGC TCCCCCTTGG TGCCCAGCTC CAGCGATTCA 567

45 GCCCAGCACG GCCCCTTCGT GAACCCCTTG GGCCTAGGTT CAGAGAGACG GCAAGGGATG 627

50 TTGTATCCCT GGAGATGGTG GTTGGAGACA TAACCGCATT TCTC 671

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 38 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

Met Arg Leu Ser Lys Thr Leu Val Asp Met Asp Met Ala Asp Tyr Ser
1 5 10 15

45 Ala Ala Leu Asp Pro Ala Tyr Thr Thr Leu Glu Phe Glu Asn Val Gln
20 25 30

50 Val Leu Thr Met Gly Asn
35

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 796 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS

(B) LOCATION: join(286..312, 316..375)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

TGGATGTTTG TACATGTGTG CTGTGTGTGC GGGTCATAGA GCACATGTGT TTGTGCATGC 60
GGACCTGTTG GAGTGCCCTG TTCTTCCTGC ATCTTTATCC TGTATGGGCG TTTTGTCTGTG 120
TGCCCATATT TGTACCTGCT GTGTATATAT GCAGTTCCTT GTGCTGCGGG CGGGGGTCAG 180
CGGTCTCTGG TGTGCACGAC TGCACAGACC CAAATGCAGG ACTCTGTTGT TGCCACTCAC 240
CAAGTGAGAT TCATATCAGC AACATGTCCG TTTGTCTCTG AGCAG ATT TGT TGC 294
Ile Cys Cys
1
CGC TGC GTC TCG CCA GAT TGA GGC ATC CCC TCC GAC ATC ACT GGA GCA 342
Arg Cys Val Ser Pro Asp Gly Ile Pro Ser Asp Ile Thr Gly Ala
5 10 15
TAT CTG GAG GGG TGG ACA GTT CTC CAC AGG GAG GTAGGGGAAA AGAGGAGGCC 395
Tyr Leu Glu Gly Trp Thr Val Leu His Arg Glu
20 25
CGGAAACCCC TCCTGGAGGG AAGAGCCCCA TCGGTCCCAG GCCAGCCTCA GAGGAGAGGG 455
GGCAGGCAGC TGGCTGAGGT CAGCCTGCCA CCCTGCTTCC TTCTGTGTCT TGGAGCCACT 515
CAGCCAGTAT GAGGCTGCAG CTCCAGCTGA GGTCTGGAAT CTTGTGGTCA GCTCAGCTAG 575
GGTGAGGAGG CAGCTGCTGG GCACTGCTTG TTGTCAGCTC AGCAGGTGCT CACCTGCCCC 635
TGCCGTCCAG TCACGTGTGA CCTTGGGCAT GTCACCTCCC CTATCCTGGC TTCTGTATCT 695
TCTACAAAAC AGGCTTCATT CCCCCAGGCC TGCTGGCTGG ACGGCTTTTA GGCCTGTCTG 755
AGGACCACGC CAGGAGCGCA AGGCAAAAAC ACACCAGAGA T 796

(2) INFORMATION FOR SEQ ID NO:37:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 29 amino acids

(B) TYPE: amino acid

(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

Ile Cys Cys Arg Cys Val Ser Pro Asp Gly Ile Pro Ser Asp Ile Thr
1 5 10 15
Gly Ala Tyr Leu Glu Gly Trp Thr Val Leu His Arg Glu
20 25

(2) INFORMATION FOR SEQ ID NO:38:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 634 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
(B) LOCATION: 326...499

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

[illegible]

(2) INFORMATION FOR SEQ ID NO:39:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 58 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Thr Arg Pro His Gln Lys Ala Pro Thr Ser Thr Arg Pro Thr Ala Trp
1 5 10 15
Val Ser Ala Pro Cys Val Pro Ser Ala Gly Thr Gly Pro Arg Ala Asn
20 25 30
Thr Thr Val Pro Arg Ala Val Thr Ala Ala Arg Ala Ser Ser Gly Gly
35 40 45
Ala Cys Gly Arg Thr Thr Cys Thr Pro Ala
50 55

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 458 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: join(171..173, 177..265)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

AGAGAGTTCA TAGCACCTTT CCAGCTCCTG GTGGGTTCAG GAGAGAACTC CCGGGATGAA 60
GAGATGAGAG CACTGAGGTT GGGGGGTCAA CTGGATAGCC AGGGCCCTAG TTCTGTCTTA 120
AGAGGAGGAA GTTGTGTCTT CTCCATCCAA CCATCCAAAG CCCTCCCCAG ATT 173
Ile
1
TAG CCG GCA GTG CGT GGT GGA CAA AGA CAA GAG GAA CCA GTG CCG CTA 221
Pro Ala Val Arg Gly Gly Gln Arg Gln Glu Glu Pro Val Pro Leu
5 10 15
CTG CAG GCT CAA GAA ATG CTT CCG GGC TGG CAT GAA GAA GGA 263
Leu Gln Ala Gln Glu Met Leu Pro Gly Trp His Glu Glu Gly
20 25 30
AGGTGAGCCT CGGCCCTCCC CGCCCCACCA CCACTGCCCC ACCTGCACCC ACAGCTCCCC 323
GACAGTCATT TACAACGTGA GCCACACTTT ATGACTCAGT GGCAGGCCCC AGGGTGAAGT 383
GCTAATGGCT GAGAAGAGGG AGGGCCTGGA AATCTGACCA TAGGGAGCGG CTGGGCTTGG 443
TCTTGAGAAA GATTC 458

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 30 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Ile Pro Ala Val Arg Gly Gly Gln Arg Gln Glu Glu Pro Val Pro Leu
1 5 10 15
Leu Gln Ala Gln Glu Met Leu Pro Gly Trp His Glu Glu Gly
20 25 30

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 662 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 84..188

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

TCCCACTCCT CATCAGTCAC AGACACCCCC ACCCCCTACT CCATCCCTGT TCTCCCTCCT 60
CACCTCTCTG TGCCTCCTCA CAG CCG TCC AGA ATG AGC GGG ACC GGA TCA 110
Pro Ser Arg Met Ser Gly Thr Gly Ser
1 5
GCA CTC GAA GGT CAA GCT ATG AGG ACA GCA GCC TGC CCT CCA TCA ATG 158
Ala Leu Glu Gly Gln Ala Met Arg Thr Ala Ala Cys Pro Pro Ser Met
10 15 20 25
CGC TCC TGC AGG CGG AGG TCC TGT CCC GAC AGGTACCGGG GTGATCCTGC 208
Arg Ser Cys Arg Arg Ser Cys Pro Asp
30 35
CACCCACCCA GGGGATCCCC CAACTACAG AGGAGCTCAC CTCCTCCACC TCCATTCTCC 268
CCAGCCAGGC CCTGGAGCAG CTGACGGGAG GGGCCTCAGA TATTACAGAA GGGACACTGA 328
GTGCGGTTTC ACATGGCCCA GTTTGCAGCA AGGGCAGGAA TCGAACCTGG CGCCCTGGGG 388
CACTTTCTAA TTCATCCTAC TGCCTGCATC CCACAGGCCA AGCAGAGTCT TCACCTTCAC 448
TGAGGGCCTG CGATCAGCTC AGCTCCGAGA GAACAGAGCA GTGGCTCAGT GGAGAGAGGT 508
GGCAAAGTGG GGCCAGCCCC TTCCCTTGCT GAGTGACCTT GGGCAAGTCA CAGCACCTCT 568
CTGAGCCATG GTTGCTCAT TGTCAGAAAA GGATGATGAT TTTTGGCCCT GCTTCTCCTC 628

TAAGGCTGAC AGACTCCTTG GGGCTCTAAA GCTG

662

5 (2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 35 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Pro Ser Arg Met Ser Gly Thr Gly Ser Ala Leu Glu Gly Gln Ala Met
1 5 10 15
Arg Thr Ala Ala Cys Pro Pro Ser Met Arg Ser Cys Arg Arg Arg Ser
20 25 30
Cys Pro Asp
35

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 647 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 185..340

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

TTCTCCCTCA TCCCTGCCTC CTCCCTCCCT CCGTTTTTAC CCTGAGCTTC CTTCAGAGCT 60
GGAGGGCACC CACTATCCAG CCCCCTCCCC ACATCTGATT CCAGGGAGGG GGCTCTGTGC 120
AGGGGACAGA GAATGCGGGA GGGCCCGGAC ATCTCCAGCA TTTTCTTCCC TGTATCTCTC 180
GAAG ATC ACC TCC CCC GTC TCC GGG ATC AAC GGC GAC ATT CGG GCG AAG 229
Ile Thr Ser Pro Val Ser Gly Ile Asn Gly Asp Ile Arg Ala Lys
1 5 10 15
AAG ATT GCC AGC ATC GCA GAT GTG TGT GAG TCC ATG AAG GAG CAG CTG 277
Lys Ile Ala Ser Ile Ala Asp Val Cys Glu Ser Met Lys Glu Gln Leu
20 25 30
CTG GTT CTC GTT GAG TGG GCC AAG TAC ATC CCA GCT TTC TGC GAG CTC 325
Leu Val Leu Val Glu Trp Ala Lys Tyr Ile Pro Ala Phe Cys Glu Leu
35 40 45
CCC CTG GAC GAC CAG GTGAGGATGG GCGTGGATGG TGGGCAGTAG TGGGCAGTGG 380
Pro Leu Asp Asp Gln

GCGGGGCAGC CAGGGGGCTG CTGGCCCACC TGGGATATAG CCGTGGACTG GCTTGATTTT 440
 5 ATTTTATTTA ACAAATATG TAGTGCACAC ACGTGTCTGA AACTTTAAAT CACCTTACAA 500
 ATATTAATC AGTTAGCTCC TCCAACAAC CTATGAGGTA GGTACTAAGG TACTATTATT 560
 10 ACTGCCATCT CATAGGTGAG AGATTGGGGC ACAGAGAGGT TAAGTAACCT GCTCAAGGTC 620
 ACATAGCTAC TATCCAGCAT AGCTGGG 647

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 52 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

Ile Thr Ser Pro Val Ser Gly Ile Asn Gly Asp Ile Arg Ala Lys Lys
 1 5 10 15
 Ile Ala Ser Ile Ala Asp Val Cys Glu Ser Met Lys Glu Gln Leu Leu
 20 25 30
 Val Leu Val Glu Trp Ala Lys Tyr Ile Pro Ala Phe Cys Glu Leu Pro
 35 40 45
 Leu Asp Asp Gln
 50

(2) INFORMATION FOR SEQ ID NO:46:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 844 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 429..515

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:46:

ATTTTACAA AGCACCCTTC ATAATTCTCC ATAGCTGGTC CATGGGTGGG AATTGTTGGAC 60
 55 CCACAGTTTT GGAACTTTTT GGGATCATAG ACCTTTTTGA GAATCTCAA AAAGAAAAA 120
 AAGCACACAG AATGTTGCTT ACAGTTTCAT CAGGCACACA GAAGAGGCC AGCACGAAGC 180
 60 AGTTTCTTGC CCAAGGACAC AGCAGTTCAA GGACAGAGTC AGCGCGAGGT CTCTCAGCTC 240

TGAGCACATG TTCTTTCCCC TTCCAGGTTT CTAGTTTAT GGGTAGTAGT TTTATGATGC 300
 CCATTTTACA GTTCAGGCAG GTAGAGGCAG AGGGGAGCAT TAAGCTGACT TGCCCAGCGT 360
 5 CACTGAGTTG GCTACGGGCA GCCTTCCCAA GGGTACAGAT GGCAAACACT GTTCCTTATC 420
 TCTTTTCTAG GTG GCC CTG CTC AGA GCC CAT GCT GGC GAG CAC CTG CTG CTC 470
 Val Ala Leu Leu Arg Ala His Ala Gly Glu His Leu Leu Leu
 1 5 10
 10 GGA GCC ACC AAG AGA TCC ATG GTG TTC AAG GAC GTG CTG CTC CTA 515
 Gly Ala Thr Lys Arg Ser Met Val Phe Lys Asp Val Leu Leu Leu
 15 20 25
 15 GGTGAGGCGG CTGCCTGCCC TGGCCAGGGC TCCAGGGAGG GTATGCCTAG CATGGCACTC 575
 ACCCAGGCAA GGAGATTAC ATGGTGGCAT GCAAGGGTGA GGGAGACTAG TCAGGAGTGG 635
 CCCTGTCCTC AGGCTTGCAT TGGAGGGCTC CAGGACTCAG TTTTCAACTG GGTACCCAC 695
 20 TCAGATGCAA GGAAATGTGG ATGCAAGTCA CCAAATTCCC AGCATTGAAG TCAGAGCACG 755
 ATCAGGGTTA TCCCTGGAAT TACCTGTGCA TCCTTTTTTC TTTTGACAGA GTCTTGCTCT 815
 25 GTCACCTCAGG CTGGAGTGCA ATGATGTGA 844

(2) INFORMATION FOR SEQ ID NO:47:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 29 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:47:

Val Ala Leu Leu Arg Ala His Ala Gly Glu His Leu Leu Leu Gly Ala
 1 5 10 15
 Thr Lys Arg Ser Met Val Phe Lys Asp Val Leu Leu Leu
 20 25

(2) INFORMATION FOR SEQ ID NO:48:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 937 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(485..529, 533..640)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

5 GCAACACTAG TATTTTAATA TAACAATGCT ATGAGGGAGC TCGATTATTT ATCCTCATCT 60
TATAGATAAG AAAACTGAGG CACAGAGAGG TTAAGTAACT TATCCAATA TAACCAGCTA 120
TCAGGGGCAG AGCCATTTAA GCAGGGCAGT GCAGTTCCAG AATCTGGTCC TTTAACCTTG 180
10 ATGCTTTGGT GCCTATCAGG TGACCTTTGA ATGTCATCGA TCTTGTGAGT CATGTTGGTA 240
AATGGAGCTT GGGTCATGTG AAAGAGGTCC TAGAAAGCCA AGTTCCAAGC TCAGCCGGAT 300
GACTCAAGGC AGCTTATCTT CTGAATCTGG GCCTCAGCTT CCTTACCTGT GAAATGGGAG 360
15 TCACCATCCC TGCAGGTCCT CCTCCCACAG GCACCAGCTA TCTTGCCAAC TTAAAAGCCA 420
AAACTAGAGG AGAGGGGTCA ACCCAAAGTG ACTTCCCATC CTCCCTCCCT CCCAACCTT 480
20 CCAG GCA ATG ACT ACA TTG TCC CTC GGC ACT GCC CGG AGC TGG CGG AGA 529
Ala Met Thr Thr Leu Ser Leu Gly Thr Ala Arg Ser Trp Arg Arg
1 5 10 15
TGA GCC GGG TGT CCA TAC GCA TCC TTG ACG AGC TGG TGC TGC CCT TCC 577
Ala Gly Cys Pro Tyr Ala Ser Leu Thr Ser Trp Cys Cys Pro Ser
20 25 30
30 AGG AGC TGC AGA TCG ATG ACA ATG AGT ATG CCT ACC TCA AAG CCA TCA 625
Arg Ser Cys Arg Ser Met Thr Met Ser Met Pro Thr Ser Lys Pro Ser
35 40 45
35 TCT TCT TTG ACC CAG GTACAGTGCA CACCTCCTAA GCCATCCCTG ACTCTCTCTC 680
Ser Ser Leu Thr Gln
50
40 CAGAACGCTC TGCCAGACTT CTCCTATTGG GTTCTGTACA CTGAGTTCAC AGCCTCATCT 740
CATGTTAACG ACAGCCAGGA GAGGCCGTTT TCATTTAACA GATGAGGCAA GTCAAGATTT 800
GAAGAGACAA TATGGCCGGG CGCAGTGGCT CACACCTGTA ATCCCATCAC TTTGGGAGGC 860
TGAGGCGGGC GGATCACCTG AGGTCAGGGG TCAAGATGAG CCTGGCTAAC ATGGAGAAAC 920
CCCATCTCTA CTAAAAA 937

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 51 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

Ala Met Thr Thr Leu Ser Leu Gly Thr Ala Arg Ser Trp Arg Arg Ala
1 5 10 15
60 Gly Cys Pro Tyr Ala Ser Leu Thr Ser Trp Cys Cys Pro Ser Arg Ser
20 25 30

Cys Arg Ser Met Thr Met Ser Met Pro Thr Ser Lys Pro Ser Ser Ser
 35 40 45

5 Leu Thr Gln
 50

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 978 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: CDS
 (B) LOCATION: join(376..387, 391..432, 436..534, 538..610)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

GTGGCTCTGC	CAACAACTGG	CTGTGCGACC	CAGGACAAGT	CCTATCTTTG	CACTGTGTCT	60
GGGTTTCCCC	GTGTGTAAGA	TGAGGCGGTT	GCTAGGTGCT	TATTGGATGC	ATTCCTCAAG	120
TCCCGCCCTC	CATCTCCTAT	TCCCCTCTCT	TCTGGTTTAG	TGCTTTAGGA	AATGTGGCAG	180
AAATCTTTTT	CTGCCTGTGT	CTAGGAAATC	ATAATTCATG	CTGGCGTACC	CTGGTTGTTG	240
AGGTCCCTGA	ATCCTTGTGC	CCACACTGCT	GAAGACTCCT	TGTGTGACAC	AAGTCAGGGG	300
ACATCTGGGT	CTTGACTCCC	CAGATGCTCC	AGGTGGACCC	TGCTGCCCTC	CCTTGCCCAC	360
CCTCTTCCAT	TGTAG	ATG CCA AGG GGC TGA GCG ATC CAG GGA AGA TCA AGC	411			
	Met Pro Arg Gly	Ala Ile Gln Gly Arg Ser Ser				
	1	5 10				
GGC TGC GTT CCC AGG TGC AGG TGA GCT TGG AGG ACT ACA TCA ACG ACC	459					
Gly Cys Val Pro Arg Cys Arg	Ala Trp Arg Thr Thr Ser Thr Thr					
15	20 25					
GCC AGT ATG ACT CGC GTG GCC GCT TTG GAG AGC TGC TGC TGC TGC TGC	507					
Ala Ser Met Thr Arg Val Ala Ala Leu Glu Ser Cys Cys Cys Cys Cys						
30	35 40					
CCA CCT TGC AGA GCA TCA CGT GGC AGA TGA TCG AGC AGA TCC AGT TCA	555					
Pro Pro Cys Arg Ala Ser Arg Gly Arg	Ser Ser Arg Ser Ser Ser					
45	50 55					
TCA AGC TCT TCG GCA TGG CCA AGA TTG ACA ACC TGT TGG AGG AGA TGC	603					
Ser Ser Ser Ser Ala Trp Pro Arg Leu Thr Thr Cys Trp Arg Arg Cys						
60	65 70					
TGC TGG GAGGTCCGTG CCAAGCCCAG GAGGGGCGGG GTTGGATTGG GGACTCCCCA	659					
Cys Trp	75					
GGAGACAGGC CTCACACAGT GAGCTCACCC CTCAGCTCCT TGGCTTCCCC ACTGTGCCGC	719					

TTTGGGCAAG TTGCTTAACC TGTCTGTGCC TCAGTTTCCT CACCAGAAAA ATGGGAACAA 779
 5 GGCAATGGTC TATTTGTTCA GGCACCGAGA ACCTAGCACG TGCCAGTCAC TGTTCCTAAGT 839
 GCTGGCAATT CAGCAAAGAA CAAGATCTTT GCCCTCGGGG AGGCTGTGTG TGTGTGATAT 899
 GTATGGATGC GTGGATATCT GTGTATATGC CCGTATGTGC GTGCATGTGT ATATAAAGCC 959
 10 TCACATTTTA TGATTTTGA 978

(2) INFORMATION FOR SEQ ID NO:51:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 75 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

Met Pro Arg Gly Ala Ile Gln Gly Arg Ser Ser Gly Cys Val Pro Arg
 1 5 10 15
 Cys Arg Ala Trp Arg Thr Thr Ser Thr Thr Ala Ser Met Thr Arg Val
 20 25 30
 Ala Ala Leu Glu Ser Cys Cys Cys Cys Cys Pro Pro Cys Arg Ala Ser
 35 40 45
 Arg Gly Arg Ser Ser Arg Ser Ser Ser Ser Ser Ser Ser Ala Trp Pro
 50 55 60
 Arg Leu Thr Thr Cys Trp Arg Arg Cys Cys Trp
 65 70 75

(2) INFORMATION FOR SEQ ID NO:52:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 984 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(443..490, 494..595)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GGGACACATA GATGCTATAA GTAGGTCAGT TGGCTGCAGC AGAGATGTGG GGGATGAGGC 60
 TGAAAGGTGA GGCGGGACCA AATGGTTGAA GGACTTGACAC TCCAAGGAGC TTTGAGAGCC 120
 60 ATTGATTACA TCCATTATGT TACTATGTGA CCAATACATT ACTCATTAGA ACATTTACGT 180

(2) INFORMATION FOR SEQ ID NO:54:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1103 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: join(289..429, 433..477, 481..492, 496..603, 607
..630, 634..750, 754..810, 814..843, 847..1023,
1027..1071, 1075..1103)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

```
TTTGGGAGAA GCAGTCCAAG TCTGCATATC AAATAAATGA TGGAGGAGAT GGGTGGTAGG      60
ACCTTCCAGA CCTCATAAAA CTTAGGCTTT ATGATCTGGG ACTCACAGAA GGTGAGCAA      120
TAAAAGACCT TAGGGATTAT CTGGCTTAAT TAATTCTCTC ATTTTATAGA GGAAGAAATT      180
AAGTCAAGGT GGGGCAGGGT GGGAGGGGAG AACTTTCCCG GGGCTCTTCA TTTACTCCCA      240
CAAAGGCTGG AATTTTGAGC AGCCCCTGTC TGTCTGTTTG TCCTTCCA GCC ACC CCT      297
                                   Ala Thr Pro
                                   1

GAG ACC CCA CAG CCC TCA CCG CCA GGT GGC TCA GGG TCT GAG CCC TAT      345
Glu Thr Pro Gln Pro Ser Pro Pro Gly Gly Ser Gly Ser Glu Pro Tyr
      5              10              15

AAG CTC CTG CCG GGA GCC GTC GCC ACA ATC GTC AAG CCC CTC TCT GCC      393
Lys Leu Leu Pro Gly Ala Val Ala Thr Ile Val Lys Pro Leu Ser Ala
      20              25              30              35

ATC CCC CAG CCG ACC ATC ACC AAG CAG GAA GTT ATC TAG CAA GCC GCT      441
Ile Pro Gln Pro Thr Ile Thr Lys Gln Glu Val Ile      Gln Ala Ala
              40              45              50

GGG GCT TGG GGG CTC CAC TGG CTC CCC CCA GCC CCC TAA GAG AGC ACC      489
Gly Ala Trp Gly Leu His Trp Leu Pro Pro Ala Pro      Glu Ser Thr
              55              60              65

TGG TGA TCA CGT GGT CAC GGC AAA GGA AGA CGT GAT GCC AGG ACC AGT      537
Trp      Ser Arg Gly His Gly Lys Gly Arg Arg Asp Ala Arg Thr Ser
              70              75              80

CCC AGA GCA GGA ATG GGA AGG ATG AAG GGC CCG AGA ACA TGG CCT AAG      585
Pro Arg Ala Gly Met Gly Arg Met Lys Gly Pro Arg Thr Trp Pro Lys
              85              90              95

GCA CAT CCC ACT GCA CCC TGA CGC CCT GCT CTG ATA ACA AGA CTT      630
Ala His Pro Thr Ala Pro      Arg Pro Ala Leu Ile Thr Arg Leu
              100              105              110

TGA CTT GGG GAG ACC CTC TAC TGC CTT GGA CAA CTT TCT CAT GTT GAA      678
```

Leu Gly Glu Thr Leu Tyr Cys Leu Gly Gln Leu Ser His Val Glu
115 120 125

5 GCC ACT GCC TTC ACC TTC ACC TTC ATC CAT GTC CAA CCC CCG ACT TCA 726
Ala Thr Ala Phe Thr Phe Thr Phe Ile His Val Gln Pro Pro Thr Ser
130 135 140

10 TCC CAA AGG ACA GCC GCC TGG AGA TGA CTT GAG CCT TAC TTA AAC CCA 774
Ser Gln Arg Thr Ala Ala Trp Arg Leu Glu Pro Tyr Leu Asn Pro
145 150 155

15 GCT CCC TTC TTC CCT AGC CTG GTG CTT CTC CTC TCC TAG CCC CGG TCA 822
Ala Pro Phe Phe Pro Ser Leu Val Leu Leu Leu Ser Pro Arg Ser
160 165 170

TGG TGT CCA GAC AGA GCC CTG TGA GGC TGG GTC CAA TTG TGG CAC TTG 870
Trp Cys Pro Asp Arg Ala Leu Gly Trp Val Gln Leu Trp His Leu
175 180 185

20 GGG CAC CTT GCT CCT CCT TCT GCT GCT GCC CCC ACC TCT GCT GCC TCC 918
Gly His Leu Ala Pro Pro Ser Ala Ala Ala Pro Thr Ser Ala Ala Ser
190 195 200

25 CTC TGC TGT CAC CTT GCT CAG CCA TCC CGT CTT CTC CAA CAC CAC CTC 966
Leu Cys Cys His Leu Ala Gln Pro Ser Arg Leu Leu Gln His His Leu
205 210 215

30 TAC AGA GGC CAA GGA GGC CTT GGA AAC GAT TCC CCC AGT CAT TCT GGG 1014
Tyr Arg Gly Gln Gly Gly Leu Gly Asn Asp Ser Pro Ser His Ser Gly
220 225 230

35 AAC ATG TTG TAA GCA CTG ACT GGG ACC AGG CAC CAG GCA GGG TCT AGA 1062
Asn Met Leu Ala Leu Thr Gly Thr Arg His Gln Ala Gly Ser Arg
235 240 245

40 AGG CTG TGG TGA GGG AAG ACG CCT TTC TCC TCC AAC CCA AC 1103
Arg Leu Trp Gly Lys Thr Pro Phe Ser Ser Asn Pro
250 255 260

(2) INFORMATION FOR SEQ ID NO:55:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 261 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

Ala Thr Pro Glu Thr Pro Gln Pro Ser Pro Pro Gly Gly Ser Gly Ser
1 5 10 15

55 Glu Pro Tyr Lys Leu Leu Pro Gly Ala Val Ala Thr Ile Val Lys Pro
20 25 30

Leu Ser Ala Ile Pro Gln Pro Thr Ile Thr Lys Gln Glu Val Ile Gln
35 40 45

60 Ala Ala Gly Ala Trp Gly Leu His Trp Leu Pro Pro Ala Pro Glu Ser

	50		55		60											
5	Thr 65	Trp	Ser	Arg	Gly	His 70	Gly	Lys	Gly	Arg	Arg 75	Asp	Ala	Arg	Thr	Ser 80
	Pro	Arg	Ala	Gly	Met 85	Gly	Arg	Met	Lys	Gly 90	Pro	Arg	Thr	Trp	Pro 95	Lys
10	Ala	His	Pro	Thr 100	Ala	Pro	Arg	Pro	Ala 105	Leu	Ile	Thr	Arg	Leu 110	Leu	Gly
	Glu	Thr	Leu 115	Tyr	Cys	Leu	Gly	Gln 120	Leu	Ser	His	Val	Glu 125	Ala	Thr	Ala
15	Phe	Thr 130	Phe	Thr	Phe	Ile	His 135	Val	Gln	Pro	Pro	Thr 140	Ser	Ser	Gln	Arg
20	Thr 145	Ala	Ala	Trp	Arg	Leu	Glu 150	Pro	Tyr	Leu	Asn 155	Pro	Ala	Pro	Phe	Phe 160
	Pro	Ser	Leu	Val 165	Leu	Leu	Ser	Pro	Arg 170	Ser	Trp	Cys	Pro	Asp 175	Arg	
25	Ala	Leu	Gly	Trp 180	Val	Gln	Leu	Trp	His 185	Leu	Gly	His	Leu	Ala 190	Pro	Pro
30	Ser	Ala	Ala	Ala 195	Pro	Thr	Ser	Ala	Ala 200	Ser	Leu	Cys	Cys 205	His	Leu	Ala
35	Gln	Pro 210	Ser	Arg	Leu	Leu	Gln 215	His	His	Leu	Tyr	Arg 220	Gly	Gln	Gly	Gly
40	Leu	Gly 225	Asn	Asp	Ser	Pro 230	Ser	His	Ser	Gly	Asn 235	Met	Leu	Ala	Leu	Thr 240
	Gly	Thr	Arg	His	Gln 245	Ala	Gly	Ser	Arg	Arg 250	Leu	Trp	Gly	Lys	Thr 255	Pro
	Phe	Ser	Ser	Asn	Pro 260											

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

GGGCACTGGG AGGAGGCAGT

20

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

5 GCCTGTAGGA CCAACCTACC 20

(2) INFORMATION FOR SEQ ID NO:58:

10 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
15 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:

TCTGGTGTGC ACGACTGCAC 20

20 (2) INFORMATION FOR SEQ ID NO:59:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
25 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:

30 CTGGAGCTGC AGCCTCATAC 20

(2) INFORMATION FOR SEQ ID NO:60:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
35 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:

40 AAGGCTCCCT TAGATGCCTG 20

(2) INFORMATION FOR SEQ ID NO:61:

50 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

55 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:

CCACTCAGGG AGAAGACAGA CCT 23

(2) INFORMATION FOR SEQ ID NO:62:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:

CCTAGTTCTG TCCTAAGAGG

20

(2) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

GTCATAAAGT GTGGCTACAG

20

(2) INFORMATION FOR SEQ ID NO:64:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

CCACCCCCTA CTCCATCCCT GT

22

(2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

CCCTCCCGTC AGCTGCTCCA

20

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

GTGCAGGGGA CAGAGAATGC

20

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 22 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

AATCAAGCCA GTCCACGGCT AT

22

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 23 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

GCCCAGCGTC ACTGAGTTGG CTA

23

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

TTGCCTGGGT GAGTGCCATG

20

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

GCACCAGCTA TCTTGCCAAC

20

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:
AGGAGAAGTC TGGCAGAGCG 20

10 (2) INFORMATION FOR SEQ ID NO:72:

- 15 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:
20 CTCCTTGTGT GACACAAGTC 20

 (2) INFORMATION FOR SEQ ID NO:73:

- 25 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:
30 CTCACTGTGT GAGGCCTGTC 20

 (2) INFORMATION FOR SEQ ID NO:74:

- 35 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
 (D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:
45 TGGTTGATTG GCCACGCCTG 20

 (2) INFORMATION FOR SEQ ID NO:75:

- 50 (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 20 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: single
55 (D) TOPOLOGY: linear

 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:
60 ATCCTGGTTC TACCTTCTAG 20

(2) INFORMATION FOR SEQ ID NO:76:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:76:

CATTACTCC CACAAAGGCT

20

(2) INFORMATION FOR SEQ ID NO:77:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:77:

GACCACGTGA TCACCAGGTG

20

(2) INFORMATION FOR SEQ ID NO:78:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1441 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 20..1414

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:78:

CTCCAAAACC CTCGTCGAC ATG GAC ATG GCC GAC TAC AGT GCT GCA CTG GAC
Met Asp Met Ala Asp Tyr Ser Ala Ala Leu Asp
1 5 10

52

CCA GCC TAC ACC ACC CTG GAA TTT GAG AAT GTG CAG GTG TTG ACG ATG
Pro Ala Tyr Thr Thr Leu Glu Phe Glu Asn Val Gln Val Leu Thr Met
15 20 25

100

GGC AAT GAC ACG TCC CCA TCA GAA GGC ACC AAC CTC AAC GCG CCC AAC
Gly Asn Asp Thr Ser Pro Ser Glu Gly Thr Asn Leu Asn Ala Pro Asn
30 35 40

148

AGC CTG GGT GTC AGC GCC CTG TGT GCC ATC TGC GGG GAC CGG GCC ACG
Ser Leu Gly Val Ser Ala Leu Cys Ala Ile Cys Gly Asp Arg Ala Thr
45 50 55

196

GGC AAA CAC TAC GGT GCC TCG AGC TGT GAC GGC TGC AAG GGC TTC TTC
Gly Lys His Tyr Gly Ala Ser Ser Cys Asp Gly Cys Lys Gly Phe Phe
60 65 70 75

244

	GAC CGC CAG TAT GAC TCG CGT GGC CGC TTT GGA GAG CTG CTG CTG CTG	1012
	Asp Arg Gln Tyr Asp Ser Arg Gly Arg Phe Gly Glu Leu Leu Leu Leu	
	320 325 330	
5	CTG CCC ACC TTG CAG AGC ATC ACC TGG CAG ATG ATC GAG CAG ATC CAG	1060
	Leu Pro Thr Leu Gln Ser Ile Thr Trp Gln Met Ile Glu Gln Ile Gln	
	335 340 345	
10	TTC ATC AAG CTC TTC GGC ATG GCC AAG ATT GAC AAC CTG TTG CAG GAG	1108
	Phe Ile Lys Leu Phe Gly Met Ala Lys Ile Asp Asn Leu Leu Gln Glu	
	350 355 360	
15	ATG CTG CTG GGA GGG TCC CCC AGC GAT GCA CCC CAT GCC CAC CAC CCC	1156
	Met Leu Leu Gly Gly Ser Pro Ser Asp Ala Pro His Ala His His Pro	
	365 370 375	
20	CTG CAC CCT CAC CTG ATG CAG GAA CAT ATG GGA ACC AAC GTC ATC GTT	1204
	Leu His Pro His Leu Met Gln Glu His Met Gly Thr Asn Val Ile Val	
	380 385 390 395	
25	GCC AAC ACA ATG CCC ACT CAC CTC AGC AAC GGA CAG ATG TGT GAG TGG	1252
	Ala Asn Thr Met Pro Thr His Leu Ser Asn Gly Gln Met Cys Glu Trp	
	400 405 410	
30	CCC CGA CCC AGG GGA CAG GCA GCC ACC CCT GAG ACC CCA CAG CCC TCA	1300
	Pro Arg Pro Arg Gly Gln Ala Ala Thr Pro Glu Thr Pro Gln Pro Ser	
	415 420 425	
35	CCG CCA GGT GCG TCA GGG TCT GAG CCC TAT AAG CTC CTG CCG GGA GCC	1348
	Pro Pro Gly Ala Ser Gly Ser Glu Pro Tyr Lys Leu Leu Pro Gly Ala	
	430 435 440	
40	GTC GCC ACA ATC GTC AAG CCC CTC TCT GCC ATC CCC CAG CCG ACC ATC	1396
	Val Ala Thr Ile Val Lys Pro Leu Ser Ala Ile Pro Gln Pro Thr Ile	
	445 450 455	
	ACC AAG CAG GAA GTT ATC TAGCAAGCCG CTGGGGCTTG GGGGCTC	1441
	Thr Lys Gln Glu Val Ile	
	460 465	

(2) INFORMATION FOR SEQ ID NO:79:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 465 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:79:

Met	Asp	Met	Ala	Asp	Tyr	Ser	Ala	Ala	Leu	Asp	Pro	Ala	Tyr	Thr	Thr
1				5					10					15	
Leu	Glu	Phe	Glu	Asn	Val	Gln	Val	Leu	Thr	Met	Gly	Asn	Asp	Thr	Ser
			20					25					30		
Pro	Ser	Glu	Gly	Thr	Asn	Leu	Asn	Ala	Pro	Asn	Ser	Leu	Gly	Val	Ser
		35					40						45		

370 375 380

Met Gln Glu His Met Gly Thr Asn Val Ile Val Ala Asn Thr Met Pro
 385 390 395 400

5 Thr His Leu Ser Asn Gly Gln Met Cys Glu Trp Pro Arg Pro Arg Gly
 405 410 415

10 Gln Ala Ala Thr Pro Glu Thr Pro Gln Pro Ser Pro Pro Gly Ala Ser
 420 425 430

Gly Ser Glu Pro Tyr Lys Leu Leu Pro Gly Ala Val Ala Thr Ile Val
 435 440 445

15 Lys Pro Leu Ser Ala Ile Pro Gln Pro Thr Ile Thr Lys Gln Glu Val
 450 455 460

Ile
 465

(2) INFORMATION FOR SEQ ID NO:80:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2329 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:80:

GGGGCCCTGA TTCACGGGCC GCTGGGGCAG GGTGGGGGT TGGGGGTGCC CACAGGGTTG 60

GCTAGTGGGG TTTTGGGGGG GCAGTGGGTG CAAGGAGTTT GGTGTGTGTC TGCCGGCCGG 120

CAGGCAAACG CAACCACGCG GTGGGGGAGG CGGCTAGCGT GGTGGACGGC CCGCGTGGCC 180

CTGTGGCAGC CGAGCCATGG TTTCTAAACT GAGCCAGCTG CAGACGGAGC TCCTGGCGGC 240

40 CCTGCTCGAG TCAGGGCTGA GCAAAGAGGC ACTGATCCAG GCACTGGGTG AGCCGGGGCC 300

CTACCTCCTG GCTGGAGAAG GCCCCCTGGA CAAGGGGGAG TCCTGCGGCG GCGGTCGAGG 360

GGAGCTGGCT GAGCTGCCCA ATGGGCTGGG GGAGACTCGG GGCTCCGAGG ACGAGACGGA 420

CGACGATGGG GAAGACTTCA CGCCACCCAT CCTCAAAGAG CTGGAGAACC TCAGCCCTGA 480

GGAGGCGGCC CACCAGAAAG CCGTGGTGGG GACCCTTCTG CAGGAGGACC CGTGGCGTGT 540

50 GGCGAAGATG GTCAAGTCCT ACCTGCAGCA GCACAACATC CCACAGCGGG AGGTGGTCGA 600

TACCACTGGC CTCAACCAGT CCCACCTGTC CCAACACCTC AACAAGGGCA CTCCCATGAA 660

GACGCAGAAG CGGGCCGCCC TGTACACCTG GTACGTCCGC AAGCAGCGAG AGGTGGCGCA 720

55 GCAGTTCACC CATGCAGGGC AGGGAGGGCT GATTGAAGAG CCCACAGGTG ATGAGCTACC 780

AACCAAGAAG GGGCGGAGGA ACCGTTTCAA GTGGGGCCCA GCATCCCAGC AGATCCTGTT 840

60 CCAGGCCTAT GAGAGGCAGA AGAACCCTAG CAAGGAGGAG CGAGAGACGC TAGTGGAGGA 900

(2) INFORMATION FOR SEQ ID NO:82:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID-NO:82:

GTAAGGCTCA AGTCATCTCC

20

(2) INFORMATION FOR SEQ ID NO:83:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:83:

Glu Gly Cys Lys Gly
1 5

(2) INFORMATION FOR SEQ ID NO:84:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:84:

Glu Gly Cys Lys Ala
1 5

(2) INFORMATION FOR SEQ ID NO:85:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:85:

Asp Gly Cys Lys Gly
1 5

(2) INFORMATION FOR SEQ ID NO:86:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..36

10 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:86:

GAC ACG TAC AGC GGC CCC CCC CCA GGC CCA GGC CCG 36
Asp Thr Tyr Ser Gly Pro Pro Pro Gly Pro Gly Pro
1 5 10

15 (2) INFORMATION FOR SEQ ID NO:87:

20 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

25 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:87:

30 Asp Thr Tyr Ser Gly Pro Pro Pro Gly Pro Gly Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:88:

35 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 36 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40 (ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..36

45 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:88:

GAC ACG TAC AGC GGC CCC CCC CCC AGG GCC AGG CCC 36
Asp Thr Tyr Ser Gly Pro Pro Pro Arg Ala Arg Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:89:

55 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 12 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

60 (ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:89:

Asp Thr Tyr Ser Gly Pro Pro Pro Arg Ala Arg Pro
1 5 10

(2) INFORMATION FOR SEQ ID NO:90:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:90:

CATGAACCCC GAAGAGTGGT G

21

(2) INFORMATION FOR SEQ ID NO:91:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:91:

GCCTCCAGAC ACCTGTTACT

20

(2) INFORMATION FOR SEQ ID NO:92:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:92:

GGCGATCATG GCAAGTTAGA AG

22

(2) INFORMATION FOR SEQ ID NO:93:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:93:

TTGGTGAGAG TATGGAAGAC C

21

(2) INFORMATION FOR SEQ ID NO:94:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:94:

GGGGTTTGCT TGTGAAACTC C

21

(2) INFORMATION FOR SEQ ID NO:95:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:95:

TTGGTGGGAA ACGGGCTTGG

20

(2) INFORMATION FOR SEQ ID NO:96:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:96:

CTCCCACTAG TACCCTAACC

20

(2) INFORMATION FOR SEQ ID NO:97:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:97:

GAGAGGGCAA AGGTCACTTC AG

22

(2) INFORMATION FOR SEQ ID NO:98:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:98:
AGTGAAGGCT ACAGACCCTA TC 22

5 (2) INFORMATION FOR SEQ ID NO:99:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
10 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:99:
15 TTCCTGGGTC TGTGTACTTG C 21

20 (2) INFORMATION FOR SEQ ID NO:100:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
25 (C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:100:
30 TGTGTTTTGG GCCAAGCACC A 21

(2) INFORMATION FOR SEQ ID NO:101:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
35 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:101:
AACCAGATAA GATCCGTGGC 20

45 (2) INFORMATION FOR SEQ ID NO:102:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
50 (B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:102:
55 AACCAGACTC ACAGCCTGAA CC 22

(2) INFORMATION FOR SEQ ID NO:103:

60 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

5 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:103:

TCACAGGGCA ATGGCTGAAC

20

10 (2) INFORMATION FOR SEQ ID NO:104:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

15

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:104:

20 TGCCGAGTCA TTGTTCAGG

20

(2) INFORMATION FOR SEQ ID NO:105:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

25

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:105:

CCTCTTATCT TATCAGCTCC AG

22

(2) INFORMATION FOR SEQ ID NO:106:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

40

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:106:

45 CTGCTCTTTG TGGTCCAAGT CC

22

(2) INFORMATION FOR SEQ ID NO:107:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

50

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:107:

60 GAGTTTGAAG GAGACCTACA G

21

(2) INFORMATION FOR SEQ ID NO:108:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:108:

ATCCACCTCT CCTTATCCCA G

21

(2) INFORMATION FOR SEQ ID NO:109:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:109:

ACTTCCGAGA AAGTTCAGAC C

21

(2) INFORMATION FOR SEQ ID NO:110:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:110:

TTTGCCTGTG TATGCACCTT G

21

(2) INFORMATION FOR SEQ ID NO:111:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:111:

GCCGAGTCCA TGCTTGCCAC

20

(2) INFORMATION FOR SEQ ID NO:112:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:112:

CTTTGCTGGT TGAGTTGGGC

20

5 (2) INFORMATION FOR SEQ ID NO:113:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 21 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

10

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:113:

15 TTCCATGACA GCTGCCCAGA G

21

(2) INFORMATION FOR SEQ ID NO:114:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:114:

TAAAGGTTGG AGCCCCCTCTG

20

(2) INFORMATION FOR SEQ ID NO:115:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

35

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:115:

40 TTGTAAGGTG ACCCCATCAG

20

(2) INFORMATION FOR SEQ ID NO:116:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

45

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:116:

55 TTGGTGATGT CCAGAAGTCC

20

(2) INFORMATION FOR SEQ ID NO:117:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid

60

(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:117:

CAGAATGTGT CAGAGTTCGC

20

(2) INFORMATION FOR SEQ ID NO:118:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:118:

CTCCCTCCTG TTCTTAAGTG

20

(2) INFORMATION FOR SEQ ID NO:119:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:119:

CTGGACTCCC AGTTCAGTCA

20

(2) INFORMATION FOR SEQ ID NO:120:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:120:

CAAGGATCCA GAAGATTGGC

20

(2) INFORMATION FOR SEQ ID NO:121:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:121:

CGTCCTCTGG GAAGATCTGC

20

(2) INFORMATION FOR SEQ ID NO:122:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:122:

GCAACAGAGC AAGACTCCAT CTCA

24

(2) INFORMATION FOR SEQ ID NO:123:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:123:

GAGTTTAATG GAAGAACTAA CC

22

(2) INFORMATION FOR SEQ ID NO:124:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 23 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:124:

CCTCATGGAG AAACATCCTA AGT

23

(2) INFORMATION FOR SEQ ID NO:125:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:125:

AGGGAGTGCA CGGCTGAGCT CCTG

24

(2) INFORMATION FOR SEQ ID NO:126:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 6254 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

(A) NAME/KEY: modified_base

(B) LOCATION: 1287..4273

(D) OTHER INFORMATION: /note= "N = A or G or C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:126:

5
10
15
20
25
30
35
40
45
50
55
60

AGCCAGCACT GTTCTTGGCA CATGGTAATC TTAACATATT TTTTCCTACA GGGAGGCCTG 60
GTGTCAGGCC GGGAGTGGGG TGGAAGGGTC CCAAAATGGA TGGAAGGGCC CCAAAATGGC 120
CGTGAGCATC CTCTGCCCTT GAGAAGAGCT AGCCCAGCTG TCTAGAGCTC CCTGCTGCTG 180
CCGCTCTCGT AAGCAGCAAG CATT TTTTGGC TCTCCTGTCT CAGCATGATG CCCCTACAAG 240
GTTCTTTTCGG GGGTGGGACC CAACGCTGCT CTCCTGATGG CCTCCCTGGC TCCCAGCACC 300
TTCCATCCCA GCTGCTCAGG GCCCCTCACC TGCCTCTCCC CCACCCTCCC CTCTGCCCCAC 360
TCCCATCGCA GGCCATAGCT CCCTGTCCCT CTCCGCTGCC ATGAGGCCTG CACTTTGCAG 420
GGCTGAAGTC CAAAGTTCAG TCCCTTCGCT AAGCACACGG ATAAATATGA ACCTTGAGAGA 480
ATTTCCCCAG CTCCAATGTA AACAGAACAG GCAGGGGGCCC TGATTACAGG GCCGCTGGGG 540
CCAGGGTTGG GGGTTGGGGG TGCCCACAGG GCTTGCTAG TGGGGTTTTG GGGGGGCAGT 600
GGGTGCAAGG AGTTTGTTTT GTGTCTGCCG GCCGGCAGGC AAACGCAACC CACGCGGTGG 660
GGGAGGCGGC TAGCGTGGTG GACCCGGGCC GCGTGGCCCT GTGGCAGCCG AGCCATGGTT 720
TCTAAACTGA GCCAGCTGCA GACGGAGCTC CTGGCGGCCC TGCTCGAGTC AGGGCTGAGC 780
AAAGAGGCAC TGATCCAGGC ACTGGGTGAG CCGGGGCCCT ACCTCCTGGC TGGAGAAGGC 840
CCCCTGGACA AGGGGGAGTC CTGCGGCGGC GGTGAGGGG AGCTGGCTGA GCTGCCCAAT 900
GGGCTGGGGG AGACTCGGGG CTCCGAGGAC GAGACGGACG ACGATGGGGA AGACTTCACG 960
CCACCCATCC TCAAAGAGCT GGAGAACCTC AGCCCTGAGG AGGCGGCCCA CCAGAAAGCC 1020
GTGGTGGAGA CCCTTCTGCA GTAAGGAGCC CTGCCCCGTC CCCGCTCCCA GGAGAGCCTA 1080
GAGGGGCCCC CCTCAGCTCC TAACGAGCCC CCCTTCTGAG TTGAGTCCCC ATGACCTTCA 1140
GCCTTTAGCC TAGTTGCTGG GAAGGGGGAC AGGGCCCATG AGAGCCCAGG GGTCCCTTGCT 1200
TGGAGGTTTG AGCCTCCAGC CCCTGAACTG CTCTCTGCA GAGTCCCAA TCCCATGAGC 1260
CCAGGCCTTT AGCCAGTCC TTGGGCNAGG GGGACATTTT CCAGGGGGTC CAAGATGGGA 1320
GAAAAAGCAG TGAATTCACA ACTCAAATGC CCACCCACCC ATCCATCCAT CCGTCCATCC 1380
ACCCATTCAT CCATTCATCC ATTCACCCAT CCATCCATCC ACATATCTTC ATCTGTGTTG 1440
TGTGTCTGTG TATCCATGTT TCTAAACCTT TATCTGTTCC AGTGTCTGTA TCCATAGGCC 1500
TGTGTCCACG TTTGTCATGT GTGTGCGTCN ACAAGTCTCT GTCCTCATGA CCATGTGTCT 1560
GTGTCCCTGT GTCCTGGCAT AAATGACCAT ACCTACCGT CCCTGAGTCT ATGTGTAGGC 1620

	CCCTGGGCTC	CATAACTGCT	TTCATGCACA	GTCCCCACCC	TCAGAGTTGA	CAAGGTTCCA	1680
5	GCACCCAGGA	CCGCAGCCCC	ACCTATGGGG	AGAGACAGCC	CTTGCTGAGC	AGATCCCGTC	1740
	CTTGCCCTCT	CCCAGGGAGG	ACCCGTGGCG	TGTGGCGAAG	ATGGTCAAGT	CCTACCTGCA	1800
	GCAGCACAAC	ATCCCACAGC	GGGAGGTGGT	CGATACCACT	GGCCTCAACC	AGTCCCACCT	1860
10	GTCCCAACAC	CTCAACAAGG	GCACTCCCAT	GAAGACGCAG	AAGCGGGCCG	CCCTGTACAC	1920
	CTGGTACGTC	CGCAAGCAGC	GAGAGGTGGC	GCAGCGTAAG	TAATGACCCT	ACCCCGCATC	1980
15	TTCCCTGGGA	GGGCCCAGGA	CTCTCCCCTA	ACTCATAGGT	GGGGGCTGGA	AGCTTCACCA	2040
	TCCCCATTAC	ACAGACAGGT	AGATGGAAAG	GAAGTCAGTG	GGATTCAACC	TGCATTTATT	2100
	ACCTATTCTG	CGCCAGGCAC	TCTGTGGGAC	GGGAGTANAC	TTGGTCCTGA	ACATCCAAAG	2160
20	ATGAATGAAA	TGGGTCCCTG	CTTTCTTTTT	CTTTTTTTAG	ATACGTGACT	CTGGAAAAAT	2220
	ATGTAAGCTC	TCTGAGCCTC	AGCTTCTTCA	TCTGTACAAT	GGGGATAGTA	AATGTGCCAA	2280
25	ATCAGAACAA	ATGCTAATGC	TTACCTGCAG	TCTTGTA CTG	AGAAGGATGG	TGAGATCATA	2340
	TCTTGGGTTG	GTAGGAAAGC	ATTCAGGGAT	TGATTAGTGA	TGTTTGCCTT	GAACACAGGT	2400
	TAAGAAAGTG	ATGGCATGTG	TGCTGTGTGT	TTGTCA TCAG	TAGATTAGAT	GATTTCTAAG	2460
30	TTCTAGCTGT	AAGCTCCTCT	GGTTCAGCGC	CATGGCAATG	AGAAAGAATC	AAGGGCAAGG	2520
	TCAGGGGAAT	GGACGAGGGA	AGGTGAGAGT	GGCCAGTACC	CCACTCACGG	CTTCTGTGTC	2580
35	CTGCAGAGTT	CACCCATGCA	GGGCAGGGAG	GGCTGATTGA	AGAGCCCACA	GGTGATGAGC	2640
	TACCAACCAA	GAAGGGGCGG	AGGAACCGTT	TCAAGTGGGG	CCCAGCATCC	CAGCAGATCC	2700
	TGTTCCAGGC	CTATGAGAGG	CAGAAGAACC	CTAGCAAGGA	GGAGCGAGAG	GTACAACGGC	2760
40	GGGCGGGAAA	CAGTGCTGGT	TTGGTCTGGG	CTGCGGCAAG	GCCAGGGGAA	GGGGAAGGTG	2820
	ACTCTAGGTC	CTGTAAAAGG	CTGTCCAGTT	GCCGAGAACT	CCTGATATTG	GCTTAGCCTG	2880
45	GCCCAGAAAA	TTGAGAATAC	TTGAACCTAA	GCCCATTCTC	CGCAGCCCCC	CTGCACCNTG	2940
	GACACCAAGC	AACCCCTTCC	ATGGATGCTC	ACCCAATTCT	ATTCTCTCTA	CAATCCTATG	3000
	GCTCTTTTGC	TCACTTTATG	AATGGAGAGA	CTGAGGTCAG	ACAGACTGTC	AATTGCCCAA	3060
50	GGTCACACAG	CAGACCTGGC	ATTGGAACCC	AGATCTGCCA	GCCTCAAACC	CTCCGGCAGA	3120
	GNTCAGCTTC	TCAGAACCCT	CCCCTTCATG	CCCAGGACAG	GGTTCCCTCT	AGCCTGGCCT	3180
55	GGAGGCTCAT	GGGTGGCTAT	TTCTGCAGGG	CGGAATGCAT	CCAGAGAGGG	GTGTCCCCAT	3240
	CACAGGCACA	GGGGCTGGGC	TCCAACCTCG	TCACGGAGGT	GCGTGTCTAC	AACTGGTTTG	3300
	CCAACCGGCG	CAAAGAAGAA	GCCTTCCGGC	ACAAGCTGGC	CATGGACACG	TACAGCGGGC	3360
60	CCCCCCCAGG	GCCAGGCCCG	GGACCTGCGC	TGCCCCGCTC	CAGCTCCCCT	GGCCTGCCTC	3420

	CACCTGCCCT CTCCCCAGT AAGGTCCACG GTAAGTGGTA TGTGGGGACA AGGGACACGT	3480
	GGGAAGGTGG GAGGGTTGGG GAGGACTGTC CCATTGACAG CAGTCACCTA AACCTCTTTG	3540
5	CACGTCAGTT TGGTTCCATT CGCAGCTGAC CCAGGGATTG GCAAAAGGTA GAAACAAAGG	3600
	CAGATTTGCT GGCTGCATAA AGGCAGACAG GCAGATGGCC TAAGCAAACC AATGGAGTTT	3660
10	GAAGTGCTGA GGGCTGTGGA GGCAGGGGAG GGCAGGGAAG TGGGGTGCTG AGGCAGGACA	3720
	CTGCTTCCCT CTCCAGGTGT GCGCTATGGA CAGCCTGCGA CCAGTGAGAC TGCAGAAGTA	3780
	CCCTCAAGCA GCGGCGGTCC CTTAGTGACA GTGTCTACAC CCCTCCACCA AGTGTCCCCC	3840
15	ACGGGCCTGG AGCCCAGCCA CAGCCTGCTG AGTACAGAAG CCAAGCTGGT GAGTGTCTTT	3900
	GCTTGTAAGG AAAACCCAAC CTCATCTTTC CTTGGCAGGG AGATTCTGGA GCAGTCCCTA	3960
20	GGGAGGCCCT GTGGGGACCC CGGCCCCCG GACACAGCTT GGCTTCCCCCT CGTAGGTCTC	4020
	AGCAGCTGGG GGCCCCCTCC CCCCTGTCAG CACCCTGACA GCACTGCACA GCTTGGAGCA	4080
	GACATCCCCA GGCCTCAACC AGCAGCCCCA GAACCTCATC ATGGCCTCAC TTCCTGGGGT	4140
25	CATGACCATC GGGCCTGGTG AGCCTGCCTC CCTGGGTCCT ACGTTCACCA ACACAGGTGC	4200
	CTCCACCCTG GTCATCGGTA AGCTGGTGGG GATGGGTGGG CACCTGGGTG GGAGGCTCAT	4260
30	GGGGCAACCG CANAATCCAG GAGCTGGAAA AGCCACTGGG ACTCATTCAT TCATTTCATC	4320
	ATTCATACAA CATGTTAGGA GAGGGGAGCA GAGAACTGAC CCCATGGCCT TTGCACTGCT	4380
	GTGGTACCCC AGGGCTCCAG GGAACCGCAG TTTGACAACT TTTGAACAAG TCACCGCTTG	4440
35	CTTTTCCCAT TAGCTTAGAC AAAGAGCTAA AGGCTCAGAG AGGGGGAATG ACTTGCCAGA	4500
	GCCACTTAAA TTAGTGGCAG GTCCCAGTGG AGGGCTGTTT CCTGACCACC TTGCCCCCTC	4560
40	TTCCAAACCA CGGGCTCTGG GAAGGAGAGG TGGTGCCCTT GGGAGGTCTT GGGCAGGGGT	4620
	GGGATATAAC TGGGGGGCCC AGCTGATTCC CTCCCCCTCC ACTCCAGGCC TGGCCTCCAC	4680
	GCAGGCACAG AGTGTGCCGG TCATCAACAG CATGGGCAGC AGCCTGACCA CCCTGCAGCC	4740
45	CGTCCAGTTC TCCCAGCCGC TGCACCCCTC CTACCAGCAG CCGCTCATGC CACCTGTGCA	4800
	GAGCCATGTG ACCCAGAACC CCTTCATGGC CACCATGGCT CAGCTGCAGA GCCCCACGG	4860
50	TGAGCACCTT GTGCCCCACA CAGCAGGAGA TGATGATAGA GGTGGCTGT CAATGGATGC	4920
	AGGGGAAAGG GGTGCCTGGC AGGCATTGCA GTCTGCATGT GTCTCTGGGA CAAGTGTGTT	4980
	TCCGTGATTG AGGGTGTCTG CAGGCCAGTG TGTTCCCATG TGAATGCACG TATCTGTGTG	5040
55	TGTGCACGAC TGCTTGTGTG AGCAGATCCC TAGTGCGTGT CTGGGTGTGT ATCGGTTGTG	5100
	CATGCATTTG TGTGCATGCC TGTGTTTCTC TGAACTCTT AGGGCCATAT GAATTTCTAA	5160
60	AATCTATTCA GACCAGTTTT GAAAATCAGC CTTGGATCTC CAACTGCTGC CCAGTCTGGC	5220
	TGTTTCAGCAG GCCCCATGCC CCCCTTTCCC CAGTCTTGAG GCCTGGGACT AGGGCTGTCA	5280

GGCACGTTTG CCACGTCTGC CCCTCTCTCC CCTGCGGCCA GCCCTCTACA GCCACAAGCC 5340
CGAGGTGGCC CAGTACACCC ACACGGGCCT GCTCCCGCAG ACTATGCTCA TCACCGACAC 5400
CACCAACCTG AGCGCCCTGG CCAGCCTCAC GCCCACCAG CAGGTAAGGT CCAGGCCTGC 5460
TGGCCCTCCC TCGGCCTGTG ACAGAGCCCC TCACCCCCAC ATCCCCCGGG CTCAGGAGGC 5520
TGCTCTGCTC CCCCAGGTCT TCACCTCAGA CACTGAGGCC TCCAGTGAGT CCGGGCTTCA 5580
CACGCCGGCA TCTCAGGCCA CCACCCTCCA CGTCCCCAGC CAGGACCCTG CCGGCATCCA 5640
GCACCTGCAG CCGGCCCACC GGCTCAGCGC CAGCCCCACA GGTGAGAGGC CCTGGCTCCA 5700
CCCCCTCCCT TACTGTCCCT GCCCCCTTCC ATGTTGGTCC CACCCCTTCT GTTGCTGTCC 5760
GTCACTGTGG GGCTGTGCAT GCAGCAGGCC TAGGGCTGCT GTGAGGAAGC ACTGGCAGGC 5820
GTGGAAGGGT GGGGTGGCTT CCATGAATCC AGTGTTTACA GTAAGATGTA CTCAGGCCAG 5880
TCCATGGGCG GCCGTGGACC CTGGCTGGGA GGCTCCCTTT GTTAAGAACC GAGGGTAGAG 5940
GTGTGACTTT GGGGTTCCCTG TTATGTGCTG TGATCCAGGA GGTGTGGCCC TGCTCCCCA 6000
TCCTGAGTAC CCCTAGGGAC AGGCAGGTGG GGTGGGTGTG GGTGCCTGGT GGGTGGCTAG 6060
CAGCCTTGTT TGCCTCTGCA GTGTCCTCCA GCAGCCTGGT GCTGTACCAG AGCTCAGACT 6120
CCAGCAATGG CCAGAGCCAC CTGCTGCCAT CCAACCACAG CGTCATCGAG ACCTTCATCT 6180
CCACCCAGAT GGCCTCTTCC TCCCAGTAAC CACGGCACCT GGGCCCTGGG GCCTGTACTG 6240
CCTGCTTGGG GGGT 6254

(2) INFORMATION FOR SEQ ID NO:127:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 631 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:127:

Met Val Ser Lys Leu Ser Gln Leu Gln Thr Glu Leu Leu Ala Ala Leu
1 5 10 15
Leu Glu Ser Gly Leu Ser Lys Glu Ala Leu Ile Gln Ala Leu Gly Glu
20 25 30
Pro Gly Pro Tyr Leu Leu Ala Gly Glu Gly Pro Leu Asp Lys Gly Glu
35 40 45
Ser Cys Gly Gly Gly Arg Gly Glu Leu Ala Glu Leu Pro Asn Gly Leu
50 55 60
Gly Glu Thr Arg Gly Ser Glu Asp Glu Thr Asp Asp Asp Gly Glu Asp
65 70 75 80

	Phe	Thr	Pro	Pro	Ile	Leu	Lys	Glu	Leu	Glu	Asn	Leu	Ser	Pro	Glu	Glu	
					85					90					95		
5	Ala	Ala	His	Gln	Lys	Ala	Val	Val	Glu	Thr	Leu	Leu	Gln	Glu	Asp	Pro	
				100					105					110			
	Trp	Arg	Val	Ala	Lys	Met	Val	Lys	Ser	Tyr	Leu	Gln	Gln	His	Asn	Ile	
			115					120					125				
10	Pro	Gln	Arg	Glu	Val	Val	Asp	Thr	Thr	Gly	Leu	Asn	Gln	Ser	His	Leu	
		130					135					140					
	Ser	Gln	His	Leu	Asn	Lys	Gly	Thr	Pro	Met	Lys	Thr	Gln	Lys	Arg	Ala	
15		145				150				155						160	
	Ala	Leu	Tyr	Thr	Trp	Tyr	Val	Arg	Lys	Gln	Arg	Glu	Val	Ala	Gln	Gln	
				165						170					175		
	Phe	Thr	His	Ala	Gly	Gln	Gly	Gly	Leu	Ile	Glu	Glu	Pro	Thr	Gly	Asp	
20				180					185					190			
	Glu	Leu	Pro	Thr	Lys	Lys	Gly	Arg	Arg	Asn	Arg	Phe	Lys	Trp	Gly	Pro	
			195					200					205				
	Ala	Ser	Gln	Gln	Ile	Leu	Phe	Gln	Ala	Tyr	Glu	Arg	Gln	Lys	Asn	Pro	
25		210					215					220					
	Ser	Lys	Glu	Glu	Arg	Glu	Thr	Leu	Val	Glu	Glu	Cys	Asn	Arg	Ala	Glu	
30		225				230					235					240	
	Cys	Ile	Gln	Arg	Gly	Val	Ser	Pro	Ser	Gln	Ala	Gln	Gly	Leu	Gly	Ser	
				245						250					255		
	Asn	Leu	Val	Thr	Glu	Val	Arg	Val	Tyr	Asn	Trp	Phe	Ala	Asn	Arg	Arg	
35				260					265					270			
	Lys	Glu	Glu	Ala	Phe	Arg	His	Lys	Leu	Ala	Met	Asp	Thr	Tyr	Ser	Gly	
		275						280					285				
40	Pro	Pro	Pro	Gly	Pro	Gly	Pro	Gly	Pro	Ala	Leu	Pro	Ala	His	Ser	Ser	
		290					295					300					
	Pro	Gly	Leu	Pro	Pro	Pro	Ala	Leu	Ser	Pro	Ser	Lys	Val	His	Gly	Val	
45		305				310					315				320		
	Arg	Tyr	Gly	Gln	Pro	Ala	Thr	Ser	Glu	Thr	Ala	Glu	Val	Pro	Ser	Ser	
				325						330					335		
	Ser	Gly	Gly	Pro	Leu	Val	Thr	Val	Ser	Thr	Pro	Leu	His	Gln	Val	Ser	
50				340					345					350			
	Pro	Thr	Gly	Leu	Glu	Pro	Ser	His	Ser	Leu	Leu	Ser	Thr	Glu	Ala	Lys	
			355					360					365				
55	Leu	Val	Ser	Ala	Ala	Gly	Gly	Pro	Leu	Pro	Pro	Val	Ser	Thr	Leu	Thr	
		370					375					380					
	Ala	Leu	His	Ser	Leu	Glu	Gln	Thr	Ser	Pro	Gly	Leu	Asn	Gln	Gln	Pro	
60		385				390					395					400	
	Gln	Asn	Leu	Ile	Met	Ala	Ser	Leu	Pro	Gly	Val	Met	Thr	Ile	Gly	Pro	

25
30
35
40
45
50
55
60

	GTCTGTCTGT	CTGTCTGCTG	AGTGAAGGCT	ACAGACCCTA	TCAAATCTAC	TCCTTTCTCT	2100
5	TTTCAGAAAT	CAACCAGACA	GTCCAGAGTT	CTGGAAATAT	GACAGACAAA	AGCAGTCAGG	2160
	ATCAGCTGCT	GTTTCTCTTT	CCAGAGTTCA	GTCAACAGAG	CCATGGGCCCT	GGGCAGTCCG	2220
	ATGATGCCTG	CTCTGAGCCC	ACCAACAAGA	AGATGCGCCG	CAACCGGTTC	AAATGGGGGC	2280
10	CCGCGTCCCA	GCAAATCTTG	TACCAGGCCT	ACGATCGGCA	AAAGAACCCC	AGCAAGGAAG	2340
	AGAGAGAGGC	CTTAGTGAG	GAATGCAACA	GGTAACACCA	CCAGAAGCTC	AGGTGGGCAG	2400
15	GTGGGCAAGT	ACACAGACCC	AGGAACCCTC	CCCTCGGTCC	TGGGATATTG	AGACACTAGT	2460
	TATACAGATA	AGTGTGGCTA	AATCAGAGCT	TCTCAAAGTA	TGTTCCACAG	TGATTGTGTG	2520
	TTTTGGGCCA	AGCACCAACA	AGTCCCCCG	CCCCCTTCA	CTCACCATCT	CCCCCTCCATC	2580
20	CATTCCCAGG	GCAGAATGTT	TGCAGCGAGG	GGTGTCCCCC	TCCAAAGCCC	ACGGCCTGGG	2640
	CTCCAACCTG	GTCAGTGGG	TCCGTGTCTA	CAACTGGTTT	GCAAACCGCA	GGAAGGAGGA	2700
25	GGCATTCGGG	CAAAAGCTGG	CCATGGACGC	CTATAGCTCC	AACCAGACTC	ACAGCCTGAA	2760
	CCCTCTGCTC	TCCCACGGCT	CCCCCACCA	CCAGCCCAGC	TCCTCTCCTC	CAAACAAGCT	2820
	GTCAGGTAAG	CAAAGGTTGG	GCCTCACTGC	CTCGGCAACC	CAACCATCCT	GGTTCTTGCC	2880
30	ACGGATCTTA	TCTGGTTTAA	GGGTTTTTCA	AGGAGCAAAC	GCTTTTGAGA	TGATCCTAGG	2940
	GCCGCTCTCT	CATTGCCAGA	ATATACTCCC	CTGGAAATAA	TGTGTGGCTC	TGATCAGTTC	3000
35	CAAGGCACTG	GGGATACATC	AGTGAACAAA	ACAAACGAGA	TAAAAATTTT	CTGCCCTCGT	3060
	GGCGCTTACA	TTCTAGAATT	AAATAGAGAA	CATGCCATAT	TTACCCTGGA	GAAAAGCAGC	3120
	CGATATTTCT	TGTGGGTGGA	CAGGGGAGGA	GAAAGCAACT	TTATTTTCTT	ATTACCCACC	3180
40	CTTGAAAACA	AGAGGTGCCG	AGTCATTGTT	CCAGGACCCT	GGTGGCACTA	ATGTTCCCTA	3240
	CTGGGTTTGT	GTTGTTTTGC	AGGAGTGCGC	TACAGCCAGC	AGGGAAACAA	TGAGATCACT	3300
45	TCCTCCTCAA	CAATCAGTCA	CCATGGCAAC	AGCGCCATGG	TGACCAGCCA	GTCGGTTTTA	3360
	CAGCAAGTCT	CCCCAGCCAG	CCTGGACCCA	GGCCACAATC	TCCTCTCACC	TGATGGTAAA	3420
	ATGGTGAGTA	CACCTGGGCC	ATTGTGCTC	TGGAGCTGAT	AAGATAAGAG	GCAAAACAAA	3480
50	CACAACCTCT	CACAAGGCCT	GCCTCAAACA	ATGAACCATT	GTAGCCCCAT	AGGGGAAAAT	3540
	GAGGGCTGTC	CAGAGTCGGA	AAGGAGAGGT	AGTGCTGGTG	ACCCACCCTT	TGGCGGGTAG	3600
55	AAAACCCAAA	GTGATGGGAT	TACAGGGGTG	AAGCACCATG	CCCAGCCAAT	AATTGTTATT	3660
	GAGTGAATGA	AGGAATGAAT	TTGAGAACTA	GTCATGCCAA	GGAATCGCTA	AGTCACATCG	3720
	TGTTGGAAAC	TGCTCTTTGT	GGTCCAAGTC	CACCCATGTT	TCTCTTGTTT	TTTTCTCTCC	3780
60	ATCAGATCTC	AGTCTCAGGA	GGAGGTTTGC	CCCCAGTCAG	CACCTTGACG	AATATCCACA	3840

5
10
15
20
25
30
35
40
45
50
55
60

GCCTCTCCCA	CCATAATCCC	CAGCAATCTC	AAAACCTCAT	CATGACACCC	CTCTCTGGAG	3900
TCATGGCAAT	TGCACAAAGT	AAGTTCTATT	CTTG GTTGA	AAACCTGGGG	GCAGGGAGAA	3960
GAAGAATGGG	AAGCAAATTA	ATGTGGTGAA	AAATAACTGT	AGGTCTCCTT	CAAAC TCACC	4020
CACAACTAGT	AAATTTGGTT	TAAC TTCTTT	AGTTTCTCAT	CTGTCTCCTT	AAATCCAATA	4080
TTTGGATTGT	TTAGCCTAAA	ACAAGAAAA	ATTGTGGAAT	GGATTTGGAT	CCTGGTCACA	4140
GTTTAGCAGC	TGTGCATCCT	GGGTCAAATC	ATTGAACCTA	TGACTCTGGG	AGACTCTCAG	4200
GCTTTAATCA	GATCTGTTTA	ATGCCCATCT	CCAACCCACA	ACTCATTTGT	GAAC TTGAGC	4260
AAGTAAATTA	ATATCTCCAA	GTCTCCGTTT	CTTTACACTT	GCCTCCCATG	GAATCTCCTA	4320
TGTAACAGGC	TCAGCCCGGT	GACTGGGACA	TTGAGCGGGG	GCTCAAATGA	TGGCATCCAT	4380
CCACCTCTCC	TTATCCCAGG	AGCTGTCTGT	GTCTTTTTCCT	CTTGCTCCCA	CAGGCCTCAA	4440
CACCTCCCAA	GCACAGAGTG	TCCCTGTCAT	CAACAGTGTG	GCCGGCAGCC	TGGCAGCCCT	4500
GCAGCCCGTC	CAGTTCTCCC	AGCAGCTGCA	CAGCCCTCAC	CAGCAGCCCC	TCATGCAGCA	4560
GAGCCCAGGC	AGCCACATGG	CCCAGCAGCC	CTTCATGGCA	GCTGTGACTC	AGCTGCAGAA	4620
CTCACACAGT	AAGGACACGG	GCATGTGGAG	GGAGGGAGCA	CTCAGGACCC	TCAGTGGCCA	4680
ACCACTTTCC	CTCTCTGGGT	CTGAAC TTTC	TCGGAAGTTT	ATTGGCTTGG	TCACTTTTCC	4740
CTGCCTATGA	TCAACCGACT	AAGACAATTT	CTCAAGCATA	ACTCTTGAGT	GTTGCTGTAC	4800
CTTTTCTAGT	CCTCTTCTCT	ACCCCTGAGA	TTCC CAGGGA	AGGGTTTGAA	TGACCTTTGC	4860
TCCCGTTCCG	TACCGGAGGC	CTCCCTGGTA	GGAAATGTGT	TCTGAGAGCA	GGTGGTTTCT	4920
CCCTCACAGC	CAAGCATCCA	CATGCTTTTCG	GGAGTTGGTT	ATGTGACTTG	GAATTTACAT	4980
GAATCTTATG	GATAACTAAT	ATGAGAAATC	CCCACTATAA	CCACCAGCCC	TTTTATCTAC	5040
CTGAGGAGAT	GGGAGCTATG	GTGTGGGATG	GGGGCTCTGT	ACCTGTGTCT	TTGCCTGTGT	5100
ATGCACCTTG	ATTCTGTCTT	CACTCTGTCT	CTCCAGTGTA	CGCACACAAG	CAGGAACCCC	5160
CCCAGTATTC	CCACACCTCC	CGGTTTCCAT	CTGCAATGGT	GGTCACAGAT	ACCAGCAGCA	5220
TCAGTACACT	CACCAACATG	TCTTCAAGTA	AACAGGTAAT	GCCAGCAGGA	TATGCGGGGG	5280
TTGGGGTGTG	GGCAGGGTGT	GATAAGGCCA	TGGATGTGCA	AAGGTTGTGG	CAAGCATGGA	5340
CTCGGCCAGA	AATTATATCC	TCTTTGCTGG	TTGAGTTGGG	CATCATCTCC	CTTAGAGAAG	5400
CCAAACTAAT	GGCCCATGAC	CCTGCCAAAT	GACACAGCTG	AGCACCTCT	CTCCTCTCTC	5460
TCTGCAGTGT	CCTCTACAAG	CCTGGTGATG	CCCACACACC	ACTTACTTCG	TGCGCAACAA	5520
CAAGGACCCT	GTTTTCCACA	CCATCACCTT	CTGGGCAGCT	GTCATGGAAA	AGCCCAGTGA	5580
CCTGACCAGC	ACCTGCGAGA	GGTCCCTGCT	ACCTGACGGA	CGTCCTGCTG	GCACCTCAGA	5640
CAATCCACTC	TCAGGAGGCG	CAGCCCGAAG	CCCAGTTTCC	CTTCTATGCA	GTATTGCCAC	5700

5
10
15
20
25
30
35
40
45
50
55
60

```

AATGCCTCTC CCACGATGTC AAGGACTCCT GTCTGTCCTG GAGGTGGGAG ACAAGGAACC      5760
ACCGAAGAGG AAGCAAGAAA GCCGTACTGT CTATGTTGTG ATCCTTCATC GAACAAACTG      5820
ATGCGAAAAC TTGAATCTGT TACTGAAATG AGGAGAGAAG GACATGTGCT ATTGAACTGA      5880
GCCAAACACA CTGTAAATAT CCACAGACTC CCTCCCCTGC CCCCATCCCA CATGATCTTG      5940
AGATTTCTTT TAAAGAAGTA AATTTGTCCA ATGGCTGTAA ACTATAAACT ACTGTAATTA      6000
AGTGCAATTT CCCCTCTGTG TCCTCTCCCC TCTGCCCTGT ATATAATACT AAAGTGTCTA      6060
TTAGTTTTCT TTGTAAAGGT CAGAGTCAAA ATTTCAAAAG TGATCTGTCC CCTCTCCCCT      6120
CATGGAGAAA CATCCTAAGT GGGGAAGTGAA GCCCCTTGTC CTCTCCCGCG GGCCTGGACA      6180
CTTATGGGGA CAGCATACCT TGGACTGACT ACCAGCTAAC TCCAGTCTCC TGACATTAAG      6240
ACACACCTCT GGATCCCTGG AGGGGCTGAA TGTAGTGTGT CAGAGTAACA TGCCAGCTTC      6300
CTGTGGGCCA GGAGCTCAGC CTGCACTCCC TAAGAAACCC CAGGGCAGGG AACTGGCTG      6360
TTTGATAGCA GAAGAAAAG TTGCAGTCTC AAAAGCCTTC CATTAAAACA ATTTATTTTA      6420
TCACTAAAAA AAA                                                    6433

```

(2) INFORMATION FOR SEQ ID NO:129:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 609 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:129:

```

Met Val Ser Lys Leu Thr Ser Leu Gln Gln Glu Leu Leu Ser Ala Leu
1          5          10          15

Leu Ser Ser Gly Val Thr Lys Glu Val Leu Val Gln Ala Leu Glu Glu
20        25        30

Leu Leu Pro Ser Pro Asn Phe Gly Val Lys Leu Glu Thr Leu Pro Leu
35        40        45

Ser Pro Gly Ser Gly Ala Glu Pro Asp Thr Lys Pro Val Phe His Thr
50        55        60

Leu Thr Asn Gly His Ala Lys Gly Arg Leu Ser Gly Asp Glu Gly Ser
65        70        75        80

Glu Asp Gly Asp Asp Tyr Asp Thr Pro Pro Ile Leu Lys Glu Leu Gln
85        90        95

Ala Leu Asn Thr Glu Glu Ala Ala Glu Gln Arg Ala Glu Val Asp Arg
100       105       110

Met Leu Ser Glu Asp Pro Trp Arg Ala Ala Lys Met Ile Lys Gly Tyr
115       120       125

```

	Met	Gln	Gln	His	Asn	Ile	Pro	Gln	Arg	Glu	Val	Val	Asp	Val	Thr	Gly	
	130						135						140				
5	Leu	Asn	Gln	Ser	His	Leu	Ser	Gln	His	Leu	Asn	Lys	Gly	Thr	Pro	Met	
	145					150					155					160	
	Lys	Thr	Gln	Lys	Arg	Ala	Ala	Leu	Tyr	Thr	Trp	Tyr	Val	Arg	Lys	Gln	
					165					170					175		
10	Arg	Glu	Ile	Leu	Arg	Gln	Phe	Asn	Gln	Thr	Val	Gln	Ser	Ser	Gly	Asn	
			180						185						190		
15	Met	Thr	Asp	Lys	Ser	Ser	Gln	Asp	Gln	Leu	Leu	Phe	Leu	Phe	Pro	Glu	
			195					200					205				
	Phe	Ser	Gln	Gln	Ser	His	Gly	Pro	Gly	Gln	Ser	Asp	Asp	Ala	Cys	Ser	
	210						215					220					
20	Glu	Pro	Thr	Asn	Lys	Lys	Met	Arg	Arg	Asn	Arg	Phe	Lys	Trp	Gly	Pro	
	225					230					235					240	
	Ala	Ser	Gln	Gln	Ile	Leu	Tyr	Gln	Ala	Tyr	Asp	Arg	Gln	Lys	Asn	Pro	
					245					250					255		
25	Ser	Lys	Glu	Glu	Arg	Glu	Ala	Leu	Val	Glu	Glu	Cys	Asn	Arg	Ala	Glu	
			260						265					270			
30	Cys	Leu	Gln	Arg	Gly	Val	Ser	Pro	Ser	Lys	Ala	His	Gly	Leu	Gly	Ser	
		275						280					285				
	Asn	Leu	Val	Thr	Glu	Val	Arg	Val	Tyr	Asn	Trp	Phe	Ala	Asn	Arg	Arg	
	290						295					300					
35	Lys	Glu	Glu	Ala	Phe	Arg	Gln	Lys	Leu	Ala	Met	Asp	Ala	Tyr	Ser	Ser	
	305					310					315					320	
	Asn	Gln	Thr	His	Ser	Leu	Asn	Pro	Leu	Leu	Ser	His	Gly	Ser	Pro	His	
					325					330					335		
40	His	Gln	Pro	Ser	Ser	Ser	Pro	Pro	Asn	Lys	Leu	Ser	Gly	Gly	Lys	Gln	
				340					345					350			
45	Arg	Leu	Gly	Leu	Thr	Ala	Ser	Ala	Thr	Gln	Pro	Ser	Trp	Phe	Leu	Pro	
		355						360					365				
	Arg	Ile	Leu	Ser	Gly	Leu	Arg	Val	Phe	Arg	Gly	Ala	Asn	Ala	Phe	Glu	
		370					375					380					
50	Met	Ile	Leu	Gly	Pro	Leu	Ser	His	Cys	Gln	Asn	Ile	Leu	Pro	Trp	Lys	
	385					390					395					400	
	Gly	Val	Arg	Tyr	Ser	Gln	Gln	Gly	Asn	Asn	Glu	Ile	Thr	Ser	Ser	Ser	
					405				410					415			
55	Thr	Ile	Ser	His	His	Gly	Asn	Ser	Ala	Met	Val	Thr	Ser	Gln	Ser	Val	
				420					425					430			
60	Leu	Gln	Gln	Val	Ser	Pro	Ala	Ser	Leu	Asp	Pro	Gly	His	Asn	Leu	Leu	
		435						440					445				

Ser Pro Asp Gly Lys Met Ile Ser Val Ser Gly Gly Gly Leu Pro Pro
450 455 460

Val Ser Thr Leu Thr Asn Ile His Ser Leu Ser His His Asn Pro Gln
465 470 475 480

Gln Ser Gln Asn Leu Ile Met Thr Pro Leu Ser Gly Val Met Ala Ile
485 490 495

Ala Gln Ser Leu Asn Thr Ser Gln Ala Gln Ser Val Pro Val Ile Asn
500 505 510

Ser Val Ala Gly Ser Leu Ala Ala Leu Gln Pro Val Gln Phe Ser Gln
515 520 525

Gln Leu His Ser Pro His Gln Gln Pro Leu Met Gln Gln Ser Pro Gly
530 535 540

Ser His Met Ala Gln Gln Pro Phe Met Ala Ala Val Thr Gln Leu Gln
545 550 555 560

Asn Ser His Met Tyr Ala His Lys Gln Glu Pro Pro Gln Tyr Ser His
565 570 575

Thr Ser Arg Phe Pro Ser Ala Met Val Val Thr Asp Thr Ser Ser Ile
580 585 590

Ser Thr Leu Thr Asn Met Ser Ser Ser Lys Gln Cys Pro Leu Gln Ala
595 600 605

Trp

(2) INFORMATION FOR SEQ ID NO:130:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 10014 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:130:

TGGGTTGCCT GTGACTGCAC TGGCGATACC CCCACAAAGC CCACTCTGAA GGTAGGAGAC 60
GGGTGGAGAG AACAGGGGG ATGGCAAGGG GGATACGAAA CAGGGAGAGG GAGGAGGGGG 120
AAGAGGATGG ACGTCTACCA GGCCCCACTT GGTGCTTGAT TTATGCCATC TCATTTCCTT 180
CTCAAACCAC CCTTTGAAGT TGATTGTACA TTTTACAGAA AAGGAAACTG AGGCTCGGAG 240
AGGAGAATCA TTTACCCAAG GTCCCAAGTA GTAGACGGTA GGTGCCTGAA TGTAATCCA 300
GGTCTCTGCC TGCTCCGGGA GGGGGTGGGG GTGAGGGAAA CAGGAGAATG TGATGGGAAA 360
ATCCGAGATG GAGCCAGCCT GGGCCAGAAA CACTGGGAGC TGTGGGAGAC GGAGAGGGGC 420
AGGGTGGGAT CACAGGGAGC AGGAGCGGGG AATTGGAGGT GAATCTGGCC CTCCCAAAC 480
TCCAGTCCAT TCTGCTCCCA GGGGAACCGG GAAACTGCGG GGGAACTGGA AGGGAGCTCC 540

	CAGAACAAGG ATCCAGAAGA TTGGCATCTG GGGCCTGGGA TTAGGTTTC TAAATCGTGG	600
5	GCCATGGGGC AGCCTTATCT CTGCAAAAGC ATTGAGGGTA GAAGTCAATG ATTTGGGAAG	660
	TTATTGAATT AGGGGATCTC GGAGGTAGGC TGTCAGTGCC TGATAGTATC AGTTAGAATG	720
	CCTGACTTGG GGTGACAATG GCTTGGAGGG GTGGGTGAGT CAAGGGTCAA ATGAGTGCCC	780
10	GTGAGTCATG ATGCCTGCCT TGTACAATTG ATAAC TGAAC ATCGGTGAGT TAGGGCCCCA	840
	GCAGTTGTAA TTAGCACCCC GGGTGT CAGC CAGAAACCAA CAAACAGCCA AATCCCTGCA	900
15	GCCCCGCCCA GCCTATCCAC CGGCGGGGGA CCGATTAACC ATTAACCCCC ACCCCTCCCC	960
	GGCAGAGCCT CCACCCCTTC ACAGAGGCTA GGCCAAGACT CCCAGCAGAT CTTCCCAGAG	1020
	GACGGTTTGA AAGGAAGGCA GAGAGGGCAC TGGGAGGAGG CAGTGGGAGG GCGGAGGGCG	1080
20	GGGGCCTTCG GGGTGGGCGC CCAGGGTAGG GCAGGTGGCC GCGGCGTGA GGCAGGGAGA	1140
	ATGCGACTCT CAAAAACCTT CGTCGACATG GACATGGCCG ACTACAGTGC TGCACTGGAC	1200
25	CCAGCCTACA CCACCCTGGA ATTTGAGAAT GTGCAGGTGT TGACGATGGG CAATGGTAGG	1260
	TGGGGGCAGA TGTGCCCAGG TGTGCCAGTG GGGGCAGGTG TGCCTGGGTC CAGGAGCAGA	1320
	TCTTTGGCAC TCAACTTTGG GGTGGGAGGA GAATGATACA AAATGGTAGG TTGGTCCTAC	1380
30	AGGCCAGCAC AGGTGTTGCC AAGTGAAGCC CATGTGCCCA GGCACAGTGA TCACAGGCAT	1440
	TCTGGGTGAA GGGAGGCCTG CAAGGGCCAA TTTCCAGCAA AAGTCGATCC CGGCTATTCC	1500
	TCCCAGGCCC TTCCAGTCCT CACTGCCTCA CAGTGGCTCT GCTTGGCGCT TGGCACAGTG	1560
35	ACATGATGGT GAGCTCCCC TTGGTGCCCA GCTCCAGCGA TTCAGCCCAG CACGGCCCCCT	1620
	TCGTGAACCC CTTGGGCCCTA GGTTCAGAGA GACGGCAAGG GATGTTGTAT CCCTGGAGAT	1680
40	GGTGGTTGGA GACATAACCG CATTTCTCGG TGTCTTTGGG ACTTTCCTAG GGAAATGAAA	1740
	TTGGCACTTA GGGAAAATGG AGCTCTCAGG GAAGTTTTGC TAACTACGAA GCCAACTCAG	1800
45	CACTGTGTGT GTTGTGTGTG CGTTCGTGTG TGATAGTGAG TTTCCATGTA GGTGTATGG	1860
	GTGGGGTGAT GCCTTCAGGA ACCCATTTGC ATATGTGTGT TCATTTGTCT CTGTGTGTGA	1920
	GTTCTGGGTC TATTTTCCTT TGTATTCATT GAGTGGGTCT GTGTTTGTGT CTAGGAGTT	1980
50	GCCCGTGT TG ATCTTGCTTA TGTATGTAAG TGTGTATGTG TGTGTACTTG TGTCTGTGGA	2040
	TGTTTGTACA TGTGTGCTGT GTGTGCGGGT CATAGAGCAC ATGCGTTTGT GCATGCGGAC	2100
	CTGTTGGAGT GCCCTGTTCT TCCTGCATCT TTATCCTGTA TGGGCGTTTT GTCGTGTGCC	2160
55	CATATTTGTA CCTGCTGTGT ATATATGCAG TTCCCTGTGC TCGGGCGGG GGTGAGCGGT	2220
	CTCTGGTGTG CACGACTGCA CAGACCCAAA TGCAGGACTC TGTTGTTGCC ACTACCAAG	2280
60	TGAGATTCAT ATCAGCAACA TGTCGGTTTG TCTCTGAGCA GATTTTGT TG CCGCTGCGTC	2340

	TCGCCAGATT GAGGCATCCC CTCCGACATC ACTGGAGCAT ATCTGGAGGG GTGGACAGTT	2400
	CTCCACAGGG AGGTAGGGGA AAAGAGGAGG CCCGAAACC CCTCCTGGAG GGAAGAGCCC	2460
5	CATCGGTCCC AGGCCAGCCT CAGAGGAGAG GGGGCAGGCA GCTGGCTGAG GTCAGCCTGC	2520
	CACCCTGCTT CCTTCTGTGT CTTGGAGCCA CTCAGCCAGT ATGAGGCTGC AGCTCCAGCT	2580
10	GAGGTCTGGA ATCTTGTGGT CAGCTCAGCT AGGGTGAGGA GGCAGCTGCT GGGCACTGCT	2640
	TGTTGTTCAGC TCAGCAGGTG CTCACCTGCC CCTGCCGTCC AGTCACGTGT GACCTTGGGC	2700
	ATGTCACCTC CCCTATCCTG GCTTCTGTAT CTTCTACAAA ACAGGCTTCA TTCCCCCAGG	2760
15	CCTGCTGGCT GGACGGCTTT TAGGCCTGTC TGAGGACCAC GCCAGGAGCG CAAGGCAAAA	2820
	ACACACCAGA GATCCCCCTG CGAGTTAGGA GGCCGGCTCC CACCCAGAA GGTGGCCAGG	2880
20	TTTTCATGCC TTCCTAGAGA AAGCTGGGGC TGGTGGCCTC CACCACAGGG AGACGCAGAC	2940
	CCTCAGAAAC AAGTCTGTGA AGTCACAACC AGCCCCAGTT TACAGATGTG AAACCTGAAGC	3000
	TCCAAAAAGT CAGGAGGTCA CTGAGTGGGG AGGTGATGGA GTGGGAACAG CCCCCAGATC	3060
25	TGGCTGAGGC CGAAGCCCTG GAGAGATCCC CGCAAGGCTC CCTTAGATGC CTGACATTCT	3120
	GCTCTTCCTG AAGCCTCACT CCCTTCTCTC CTGGCGCAGA CACGTCCCCA TCAGAAGGCA	3180
30	CCAACCTCAA CGCGCCCAAC AGCCTGGGTG TCAGCGCCCT GTGTGCCATC TGCGGGGACC	3240
	GGGCCACGGG CAAACACTAC GGTGCCTCGA GCTGTGACGG CTGCAAGGGC TTCTTCCGGA	3300
	GGAGCGTGCG GAAGAACCAC ATGTACTCCT GCAGGTGAGG AGCCTCAATT TCTTCAGCTG	3360
35	GGAAATGGGC ACACTTGGGC TCATGGCCCC AAGGTCTGTC TTCTCCCTGA GTGGGTAGGT	3420
	CCCAGAGACA GCTGCCCTTC AGGCCTTCA AGGCTCTTCT GGTTTTGTAA AAGACTTTGT	3480
40	GAATCCAAGA AGAGCATCTA TTCTAGGAAC CACATTTACT GATCATCAAG CTACTGGCTG	3540
	CCGTTTATTG AGCTCTTATC ATATGCCAGG CACAATACTA AGTCTTTGTG TGTATTTACG	3600
	TACTCCAGAG GTCAAGGTC CCAACTCAGC TCTAACACCA ACCAGCAGAG CGACCCAGGA	3660
45	CCACATGTTG CCTCTCTGAG CCTCAGTTTT CCCATGTTTA GCAGGACAGG ACTGGGCTCT	3720
	TAGAGAGTTC ATAGCACCTT TCCAGCTCCT GGTGGGTTC AAGAGAGAACT CCCGGGATGA	3780
50	AGAGATGAGA GCACTGAGGT TGGGGGGTCA ACTGGATAGC CAGGGCCCTA GTTCTGTCCT	3840
	AAGAGGAGGA AGTTGTGTCT TCTCCATCCA ACCATCCAAA GCCCTCCCCA GATTTAGCCG	3900
	GCAGTGCGTG GTGGACAAAG ACAAGAGGAA CCAGTGCCGC TACTGCAGGC TCAAGAAATG	3960
55	CTTCCGGGCT GGCATGAAGA AGGAAGGTGA GCCTCGGCCC TCCCCGCCCC ACCACCACTG	4020
	CCCCACCTGC ACCCACAGCT CCCCAGACAGT CATTTACAAC TGTAGCCACA CTTTATGACT	4080
60	CAGTGGCAGG CCCCAGGGTG ACTGGCTAAT GGCTGAGAAG AGGGAGGGCC TGGAAATCTG	4140
	ACCATAGGGA GCGGCTGGGC TTGGTCTTGA GAAAGATTCT CCCACTCCTC ATCAGTCACA	4200

	GACACCCCCA	CCCCCTACTC	CATCCCTGTT	CTCCCTCCTC	ACCTCTCTGT	GCCTCCTCAC	4260
5	CCGTCCAGAA	TGAGCGGGAC	CGGATCAGCA	CTCGAAGGTC	AAGCTATGAG	GACAGCAGCC	4320
	TGCCCTCCAT	CAATGCGCTC	CTGCAGGCGG	AGGTCCTGTC	CCGACAGGTA	CCGGGGTGAT	4380
	CCTGCCACCC	ACCCAGGGAT	CCCCCACACT	ACAGAGGAGC	TCACCTCCTC	CACCTCCATT	4440
10	CTCCCCAGCC	AGGCCCTGGA	GCAGCTGACG	GGAGGGGCCT	CAGATATTAC	AGAAGGGACA	4500
	CTGAGTGCGG	TTTCACATGG	CCCAGTTTGC	AGCAAGGGCA	GGAATCGAAC	CTGGCGCCCT	4560
15	GGGGCACTTT	CTAATTCATC	CTACTGCCTG	CATCCCACAG	GCCAAGCAGA	GTCTTCACCT	4620
	TCACTGAGGG	CCTGCGATCA	GCTCAGCTCC	GAGAGAACAG	AGCAGTGGCT	CAGTGGAGAG	4680
	AGGTGGCAAA	GTGGGGCCCA	GCCCTTCCCT	TGCTGAGTGA	CCTTGGGCAA	GTCACAGCAC	4740
20	CTCTCTGAGC	CATGGTTGCC	TCATTGTCAG	AAAAGGATGA	TGATTTTTTG	CCCTGCTTCT	4800
	CCTCTAAGGC	TGACAGACTC	CTTGGGGCTC	TAAAGCTGTT	CTCCCTCATC	CCTGCCTCCT	4860
25	CCCTCCCTCC	GTTTTTACCC	TGAGCTTCCT	TCAGAGCTGG	AGGGCACCCA	CTATCCAGCC	4920
	CCCTCCCCAC	ATCTGATTCC	AGGGAGGGGG	CTCTGTGCAG	GGGACAGAGA	ATGCGGGAGG	4980
30	GCCCGGACAT	CTCCAGCATT	TTCTTCCCTG	TATCTCTCGA	AGATCACCTC	CCCCGTCTCC	5040
	GGGATCAACG	GCGACATTCG	GGCGAAGAAG	ATTGCCAGCA	TCGCAGATGT	GTGTGAGTCC	5100
	ATGAAGGAGC	AGCTGCTGGT	TCTCGTTGAG	TGGGCCAAGT	ACATCCCAGC	TTTCTGCGAG	5160
35	CTCCCCCTGG	ACGACCAGGT	GAGGATGGGC	GTGGATGGTG	GGCAGTAGTG	GGCAGTGGGC	5220
	GGGGCAGCCA	GGGGGCTGCT	GGCCCACCTG	GGATATAGCC	GTGGACTGGC	TTGATTTTAT	5280
	TTTATTTAAC	AAAATATGTA	GTGCACACAC	GTGTCTGAAA	CTTTAAATCA	CCTTACAAAT	5340
40	ATTAACTCAG	TTAGCTCCTC	CAACAACTCT	ATGAGGTAGG	TACTAAGGTA	CTATTATTAC	5400
	TGCCATCTCA	TAGGTGAGGA	GATTGGGGCA	CAGAGAGGTT	AAGTAACCTG	CTCAAGGTCA	5460
45	CATAGCTACT	ATCCAGCATA	GCTGGGATTT	TTACAAAGCA	CCCTTCATAA	TTCTCCATAG	5520
	CTGGTCCATG	GGTGGAATT	TGGGACCCAC	AGTTTGGGAA	CTTTTGGGA	TCATAGACCT	5580
	TTTTGAGAAT	CTCAAAAAG	AAAAAAAAG	CACACAGAAT	GTTGCTTACA	GTTTCATCAG	5640
50	GCACACAGAA	GAGGCCCAGC	ACGAAGCAGT	TTCTTGCCCA	AGGACACAGC	AGTTCAAGGA	5700
	CAGAGTCAGC	GCGAGGTCTC	TCAGCTCTGA	GCACATGTTT	TTTCCCCTTC	CAGGTTTCTA	5760
55	GTTTTATGGG	TAGTAGTTTT	ATGATGCCCA	TTTCACAGTT	CAGGCAGGTA	GAGGCAGAGG	5820
	GGAGCATTAA	GCTGACTTGC	CCAGCGTCAC	TGAGTTGGCT	ACGGGCAGCC	TTCCAAGGG	5880
	TACAGATGGC	AAACACTGTT	CCTTCTCTCT	TTCAGGTGGC	CCTGCTCAGA	GCCCATGCTG	5940
60	GCGAGCACCT	GCTGCTCGGA	GCCACCAAGA	GATCCATGGT	GTTCAAGGAC	GTGCTGCTCC	6000

	TAGGTGAGGC	GGCTGCCTGC	CCTGGCCAGG	GCTCCAGGGA	GGGTATGCCT	AGCATGGCAC	6060
	TCACCCAGGC	AAGGAGATTC	ACATGGTGGC	ATGCAAGGGT	GAGGGAGACT	AGTCAGGAGT	6120
5	GGCCCTGTCC	TCAGGCTTGC	ATTGGAGGGC	TCCAGGACTC	AGTTTTCAAC	TGGGTACCCC	6180
	ACTCAGATGC	AAGGAAATGT	GGATGCAAGT	CACCAAATTC	CCAGCATTGA	AGTCAGAGCA	6240
10	CGATCAGGGT	TATCCCTGGA	ATTACCTGTG	CATCCTTTTT	TCTTTTGACA	GAGTCTTGCT	6300
	CTGTCACTCA	GGCTGGAGTG	CAATGATGTG	AGCAAACACT	ACCTATTTTA	ATATAACAAT	6360
	GCTATGAGGG	AGCTCGATTA	TTTATCCTCA	TCTTATAGAT	AAGAAAAC TG	AGGCACAGAG	6420
15	AGGTTAAGTA	ACTTATCCAA	CTATAACCAG	CTATCAGGGG	CAGAGCCATT	TAAGCAGGGC	6480
	AGTGCAGTTC	CAGAACTGG	TCCTTTAACC	TTGATGCTTT	GGTGCCTATC	AGGTGACCTT	6540
	TGAATGTCAT	CGATCTTG TG	AGTCATGTTG	GTAAATGGAG	CTTGGGTCAT	GTGAAAGAGG	6600
20	TCCTAGAAAG	CCAAGTTCCA	AGCTCAGCCG	GATGACTCAA	GGCAGCTTAT	CTTCTGAATC	6660
	TGGGCCTCAG	CTTCCTTACC	TGTGAAATGG	GAGTCACCAT	CCCTGCAGGT	CCTCCTCCCA	6720
25	CAGGCACCAG	CTATCTTGCC	AACTTAAAAG	CCAAAAC TAG	AGGAGAGGGG	TCAACCCAAG	6780
	GTGACTTCCC	ATCCTCCCTC	CCTCCCAACC	CTTCCAGGCA	ATGACTACAT	TGTCCCTCGG	6840
	CACTGCCCCG	AGCTGGCGGA	GATGAGCCGG	GTGTCCATAC	GCATCCTTGA	CGAGCTGGTG	6900
30	CTGCCCTTCC	AGGAGCTGCA	GATCGATGAC	AATGAGTATG	CCTACCTCAA	AGCCATCATC	6960
	TTCTTTGACC	CAGGTACAGT	GCACACCTCC	TAAGCCATCC	CTGACTCTCT	CTCCAGAACG	7020
35	CTCTGCCAGA	CTTCTCCTAT	TGGGTTCTGT	ACACTGAGTT	CACAGCCTCA	TCTCATGTTA	7080
	ACGACAGCCA	GGAGAGGCCG	TTTTCATTTA	ACAGATGAGG	CAAGTCAAGA	TTTGAAGAGA	7140
	CAATATGGCC	GGGCGCAGTG	GCTCACACCT	GTAATCCCAT	CAC TTTGGGA	GGCTGAGGCG	7200
40	GGCGGATCAC	CTGAGGTCAG	GGGTCAAGAT	GAGCCTGGCT	AACATGGAGA	AACCCCATCT	7260
	CTACTTAAAA	GTGGCTCTGC	CAACAACTGG	CTGTGCGACC	CAGGACAAGT	CCTATCTTTG	7320
45	CACTGTGTCT	GGGTTTCCCC	GTGTGTAAGA	TGAGGCGGTT	GCTAGGTGCT	TATTGGATGC	7380
	ATTCC TCAAG	TCCCGCCCTC	CATCTCCTAT	TCCCCTCTCT	TCTGGTTTAG	TGCTTTAGGA	7440
	AATGTGGCAG	AAATCTTTTT	CTGCCTGTGT	CTAGGAAATC	ATAATTCATG	CTGGCGTACC	7500
50	CTGGTTGTTG	AGGTCCCTGA	ATCCTTG TGC	CCACACTGCT	GAAGACTCCT	TGTGTGACAC	7560
	AAGTCAGGGG	ACATCTGGGT	CTTGACTCCC	CAGATGCTCC	AGCTGGACCC	TGCTGCCCTC	7620
55	CCTTGCCAC	CCTCTTCCAT	TGTAGATGCC	AAGGGGCTGA	GCGATCCAGG	GAAGATCAAG	7680
	CGGCTGCGTT	CCCAGGTGCA	GGTGAGCTTG	GAGGACTACA	TCAACGACCG	CCAGTATGAC	7740
	TCGCGTG GCC	GCTTTGGAGA	GCTGCTGCTG	CTGCTGCCCA	CCTTG CAGAG	CATCACCTGG	7800
60	CAGATGATCG	AGCAGATCCA	GTCATCAAG	CTCTTCGGCA	TGGCCAAGAT	TGACAACCTG	7860

ACCAAGCAGG AAGTTATCTA GCAAGCCGCT GGGGCTTGGG GGCTCCACTG GCTCCCCCA 9720
 GCCCCCTAAG AGAGCACCTG GTGATCACGT GGTCACGGCA AAGGAAGACG TGATGCCAGG 9780
 5 ACCAGTCCCA GAGCAGGAAT GGAAGGATG AAGGGCCCCG GAACATGGCC TAAGGCACAT 9840
 CCCACTGCAC CCTGACGCCC TGCTCTGATA ACAAGACTTT GACTTGGGGA GACCCTCTAC 9900
 TGCCTTGGAC AACTTTCTCA TGTGAAGCC ACTGCCTTCA CCTTCACCTT CATCCATGTC 9960
 10 CAACCCCCGA CTTTCATCCCA AAGGACAGCC GCCTGGAGAT GACTTGAGCC TTAC 10014

(2) INFORMATION FOR SEQ ID NO:131:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 567 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:131:

Met	Arg	Leu	Ser	Lys	Thr	Leu	Val	Asp	Met	Asp	Met	Ala	Asp	Tyr	Ser	1	5	10	15
Ala	Ala	Leu	Asp	Pro	Ala	Tyr	Thr	Thr	Leu	Glu	Phe	Glu	Asn	Val	Gln	20	25	30	
Val	Leu	Thr	Met	Gly	Asn	Gly	Pro	Ser	Ser	Pro	His	Cys	Leu	Thr	Val	35	40	45	
Ala	Leu	Leu	Gly	Ala	Trp	His	Ser	Asp	Met	Met	Ile	Leu	Leu	Pro	Leu	50	55	60	
Arg	Leu	Ala	Arg	Leu	Arg	His	Pro	Leu	Arg	His	His	Trp	Ser	Ile	Ser	65	70	75	80
Gly	Gly	Val	Asp	Ser	Ser	Pro	Gln	Gly	Asp	Thr	Ser	Pro	Ser	Glu	Gly	85	90	95	
Thr	Asn	Leu	Asn	Ala	Pro	Asn	Ser	Leu	Gly	Val	Ser	Ala	Leu	Cys	Ala	100	105	110	
Ile	Cys	Gly	Asp	Arg	Ala	Thr	Gly	Lys	His	Tyr	Gly	Ala	Ser	Ser	Cys	115	120	125	
Asp	Gly	Cys	Lys	Gly	Phe	Phe	Arg	Arg	Ser	Val	Arg	Lys	Asn	His	Met	130	135	140	
Tyr	Ser	Cys	Arg	Phe	Ser	Arg	Gln	Cys	Val	Val	Asp	Lys	Asp	Lys	Arg	145	150	155	160
Asn	Gln	Cys	Arg	Tyr	Cys	Arg	Leu	Lys	Lys	Cys	Phe	Arg	Ala	Gly	Met	165	170	175	
Lys	Lys	Glu	Ala	Val	Gln	Asn	Glu	Arg	Asp	Arg	Ile	Ser	Thr	Arg	Arg	180	185	190	
Ser	Ser	Tyr	Glu	Asp	Ser	Ser	Leu	Phe	Ser	Ile	Asn	Ala	Leu	Leu	Gln	195	200	205	

Ser Pro Pro Gly Gly Ser Gly Ser Glu Pro Tyr Lys Leu Leu Pro Gly
530 535 540

5 Ala Val Ala Thr Ile Val Lys Pro Leu Ser Ala Ile Pro Gln Pro Thr
545 550 555 560

Ile Thr Lys Gln Glu Val Ile
565

10

(2) INFORMATION FOR SEQ ID NO:132:

(i) SEQUENCE CHARACTERISTICS:
15 (A) LENGTH: 470 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

20 (xi) SEQUENCE DESCRIPTION: SEQ ID NO:132:

AAGTAAGCCT TGTTTTTCCA CACTCATTCT CCCAGGTTTT CTTTGGATAG GCTTACTTTT 60
CCATGCTGGA GGAGGGGCTA TCCCTTCATT TTGCCTCTCC CGCTTCCCTC CCTCTCCCCC 120
25 TCCCCCTGCT TTCTCTCCCT CTGCACTTTG TGAAGTCTG CTGCAGTGCT GAAGTCCAAA 180
GTTTCAAGTAAC TTGCTAAGCA CACAGATAAA TATGAACCTT GGAGAATTTA CCAATGTAAA 240
30 CAGATAGCCA AGGGTCCCTT TATCAGCACT GGCTCAGGAC AGTCGTGGGG GGTCTGAAGT 300
GGCTCAATTT TGTATTTTGT TTTTTTTGGG GGGGTGTAAA GGCGGGAGGC TGCCTGTGTC 360
CCGCTGCTGA CAGTCGGGCG TGTTACCTCG GGAACATGGT GTAGGGAAGC TGGAAGCAGG 420
35 ATAACGTGGA ACTCAACCCA AGAAACGCCA GCCTGAAGAC CATGGTCTCG 470

(2) INFORMATION FOR SEQ ID NO:133:

40 (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 467 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
45 (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:133:

TCACAGCTAT TAGCTCATCG CTGCCAAATT GCCCCTTTAC CTAGGCTTGT GTCACTTTCA 60
50 CCTTCTCATT CTCTTACTTT TACATCTCTC CTTGATATTT TGCTTTTTTCA ACTTTTGGAA 120
ATTTCTTTCT CTCTTCTACC CCTCCTCATA TTCTCTGCA CTCCCCCTC TCTAACTCAT 180
GCACTTTGTG GGGTCCAAAG TTCAGTAACT TGCAAAGCAC AGGGATAAAG ATGAACCTTG 240
55 GAAGATTTAC TCTGCTCTGA TGTAACAGA GAGTGACAAG GGTCCCTTAT CTATGTCTCA 300
GAGAAGCCTG TCCGGGGGGT GACCACTTGC TGTTGTGGC TGCACAGTGT GTTTTTTTGG 360
60 GGGGGAGGAG GAAACAGAAG GTGGGTAGAG CATGGACTCC CGCCCGCTGA TCCGTGTTAC 420

AGCCGCAGAT GGTGAGGCAG TAGAAGGCAA CAGACAGGAT GCGCTCT

467

(2) INFORMATION FOR SEQ ID NO:134:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 479 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:134:

TTTCGGGGGT GGGACCCAAC GCTGCTCTCC TGATGGCCTC CCTGGCTCCC AGCACCTTCC 60
ATCCCAGCTG CTCAGGGCCC CTCACCTGCG CCTCCCCCAC CCTCCCCTCT GCCCACTCCC 120
ATCGCAGGCC ATAGCTCCCT GTCCCTCTCC GCTGCCATGA GGCCTGCACT TTGCAGGGCT 180
GAAGTCCAAA GTTCAGTCCC TTCGCTAAGC ACACGGATAA ATATGAACCT TGGAGAATTT 240
CCCCAGCTCC AATGTAAACA GAACAGGCAG GGGCCCTGAT TCACGGGCCG CTGGGGCCAG 300
GGTTGGGGGT TGGGGGTGCC CACAGGGCTT GGCTAGTGGG GTTTTGGGGG GGCAGTGGGT 360
GCAAGGAGTT TGGTTTGTGT CTGCCGGCCG GCAGGCCAAAC GCAACCCACG CGGTGGGGGA 420
GGCGGCTAGC GTGGTGGACC CGGGCCGCGT GGCCCTGTGG CAGCCGAGCC ATGGTTTCT 479

(2) INFORMATION FOR SEQ ID NO:135:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 605 base pairs

(B) TYPE: nucleic acid

(C) STRANDEDNESS: single

(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:135:

TGGGGCCTGG GATTTAGGTT TCTAAATCGT GGGCCATGGG GCAGCCTTAT CTCTGCAAAA 60
GCATTGAGGG TAGAAGTCAA TGATTTGGGA AGTTATTGAA TTAGGGGATC TCGGAGGTAG 120
GCTGTCAGTG CCTGATAGTA TCAGTTAGAA TGCCTGACTT GGGGTGACAA TGGCTTGGAG 180
GGGTGGGTGA GTCAAGGGTC AAATGAGTGC CCGTGAGTCA TGATGCCTGC CTTGTACAAT 240
TGATAACTGA ACATCGGTGA GTTAGGGCCC CAGCAGTTGT AATTAGCACC CCGGGTGTCA 300
GCCAGAAACC AACAAACAGC CAAATCCCTG CAGCCCCGCC CAGCCTATCC ACCGGCGGGG 360
GACCGATTAA CCATTAACCC CCACCCCTCC CCGGCAGAGC CTCCACCCCT TCACAGAGGC 420
TAGGCCAAGA CTCCCAGCAG ATCTTCCAG AGGACGGTTT GAAAGGAAGG CAGAGAGGGC 480
ACTGGGAGGA GGCAGTGGGA GGGCGGAGGG CGGGGCCTT CGGGGTGGGC GCCCAGGGTA 540
GGGCAGGTGG CCGCGGCGTG GAGGCAGGGA GAATGCGACT CTCCAAAACC CTCGTCGACG 600
ACATG 605

(2) INFORMATION FOR SEQ ID NO:136:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 478 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:136:

TCCTGGAGAG	TGGGACCCAG	CGCCGCACCC	AGAGGCCTCC	TGGCTCCTGC	TGCCTCTAGC	60
CCTGCGCCCC	TGGCCCCTCT	CCACCTCCCC	CACCCTCCCT	TCTGCTCACT	CCCAATTGCA	120
GGCCATGACT	CCGGTCCGCG	TCCCTCTCAC	CCCCATGAGG	CCTGCACTTG	CAAGGCTGAA	180
GTCCAAAGTT	CAGTCCCTTC	GCTAAGCGCA	CGGATAAATA	TGAACCTTGG	AGAATTTCCC	240
CAGCTCCAAT	GTAAACAGAG	CAGGCAGGGG	CCCTGATTCA	CTGGCCGCTG	GGGCCAGGGT	300
TGGGGGCTGG	GGGTGCCCAC	AGAGCTTGAC	TAGTGGGATT	TGGGGGGGCA	GTGGGTGCAG	360
CGAGCCCGGT	CCGTTGACTG	CCAGCCTGCC	GGCAGGTAGA	CACCGGCCGT	GGGTGGGGGA	420
GGCGGCTAGC	TCAGTGGCCT	TGGGCCGCGT	GGCTGGTGGC	AGCGGAGCCA	TGGTTTCT	478

(2) INFORMATION FOR SEQ ID NO:137:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 622 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:137:

TGGGCTTGGG	TGTTAGGTTT	CCAGTTCAAG	CGACCCAGGA	CAGCTTTATC	TCAAATTGAG	60
GATAGAAGTC	AATGATCTGG	GACGTGATTG	GCTTAGGGCT	TCATAGTGGT	AGGCTTGCCA	120
GTGTCTAAAC	ATGTCAGCTG	GGTTGTCCAC	CTTGGTGAGA	CTTGGGGGCT	GCTGAGGCAA	180
GGGGTCCAAC	CAATGCCAGT	CCTGTTGGGT	GCCTGCCTTG	GAAGATTGGT	AAGTGACTAT	240
TAATGAGCGG	GAGGTGGGGG	GGGGGCAACA	GTTGTAATTA	GCACCCAGG	TGTCAGTCAG	300
AAACCAACAA	ACAGCCAAAT	CCTCGTGGCT	CCACCCAGCC	TACCCAGCAA	CGGGGGTGAT	360
TAACCATTAA	CTCCTACCCC	TCCCCACAGA	GCCTCCACCC	TCTGCAGAGG	CTAGGCCAGG	420
ACGCCAGGCT	GAGTCTCCCA	GAGGACAGTT	TGAAAGAGAG	GAAGGCAGAG	AAGGGACCTG	480
GGAGGAGGCA	GGAGGAGGGC	GGGGACGGGG	GGGGCTGGGG	CTCAGCCCAG	GGGCTTGGGT	540
GGCATCCTGG	GCCGGGCAGG	ACAGGGGGCT	AAGGCGTGGG	TAGGGGAGAA	TGCGACTCTC	600
TAAAACCCCTT	GCCGGCGATA	TG				622

(2) INFORMATION FOR SEQ ID NO:138:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 470 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:138:

TCTTGGGCAG TGGGACCAGC GCTGCTCCCA GAGGCCTCCT GGCTCCTGGT GCCTCTCTCC 60
CTGCGCCCCT GGTTCCTCGT CCACCTCCCC CACCCGCCCT TCTGCTCACT CCCAATTGCA 120
AGCCATGGCT CCCGGTCCGG TCCCTCTCGC TGCTGTGAGG CCTGCACTTG CAAGGCTGAA 180
GTCCAAAGTT CAGTCCCTTC GCTAAGCACA CGGATAAATA TGAACCTTGG AGAATTTCCC 240
CAGCTCCAAT GTAAACAGAG CAGCAGGGGG CCCTGATTCA CTAGCCGCTG GGGCCAGGGT 300
TGGGGGTTGG GGGTGCCAC AGGGCTTGAC TAGTGGGATT TGGGGGAGCA GTGGGTGCAG 360
CGAGCCTGGT CCGTTGACTG CCAGCAGTAG ACACCGGCCG TGTGTGGGGG AGGCGGCTAG 420
CTCAGTGGCC TTGGGCCGCG TGGCCTGGCG GTAGAGGAGC CATGGTTTCT 470

(2) INFORMATION FOR SEQ ID NO:139:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 557 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS:
- (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:139:

Met Val Ser Lys Leu Thr Ser Leu Gln Gln Glu Leu Leu Ser Ala Leu
1 5 10 15
Leu Ser Ser Gly Val Thr Lys Glu Val Leu Val Gln Ala Leu Glu Glu
20 25 30
Leu Leu Pro Ser Pro Asn Phe Gly Val Lys Leu Glu Thr Leu Pro Leu
35 40 45
Ser Pro Gly Ser Gly Ala Glu Pro Asp Thr Lys Pro Val Phe His Thr
50 55 60
Leu Thr Asn Gly His Ala Lys Gly Arg Leu Ser Gly Asp Glu Gly Ser
65 70 75 80
Glu Asp Gly Asp Asp Tyr Asp Thr Pro Pro Ile Leu Lys Glu Leu Gln
85 90 95
Ala Leu Asn Thr Glu Glu Ala Ala Glu Gln Arg Ala Glu Val Asp Arg
100 105 110
Met Leu Ser Glu Asp Pro Trp Arg Ala Ala Lys Met Ile Lys Gly Tyr
115 120 125

Asn Thr Ser Gln Ala Gln Ser Val Pro Val Ile Asn Ser Val Ala Gly
 450 455 460
 Ser Leu Ala Ala Leu Gln Pro Val Gln Phe Ser Gln Gln Leu His Ser
 465 470 475 480
 Pro His Gln Gln Pro Leu Met Gln Gln Ser Pro Gly Ser His Met Ala
 485 490 495
 Gln Gln Pro Phe Met Ala Ala Val Thr Gln Leu Gln Asn Ser His Met
 500 505 510
 Tyr Ala His Lys Gln Glu Pro Pro Gln Tyr Ser His Thr Ser Arg Phe
 515 520 525
 Pro Ser Ala Met Val Val Thr Asp Thr Ser Ser Ile Ser Thr Leu Thr
 530 535 540
 Asn Met Ser Ser Ser Lys Gln Cys Pro Leu Gln Ala Trp
 545 550 555

(2) INFORMATION FOR SEQ ID NO:140:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 516 amino acids
 - (B) TYPE: amino acid
 - (C) STRANDEDNESS:
 - (D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:140:

Met Asp Met Ala Asp Tyr Ser Ala Ala Leu Asp Pro Ala Tyr Thr Thr
 1 5 10 15
 Leu Glu Phe Glu Asn Val Gln Val Leu Thr Met Gly Asn Gly Pro Ser
 20 25 30
 Ser Pro His Cys Leu Thr Val Ala Leu Leu Gly Ala Trp His Ser Asp
 35 40 45
 Met Met Ile Leu Leu Pro Leu Arg Leu Ala Arg Leu Arg His Pro Leu
 50 55 60
 Arg His His Trp Ser Ile Ser Gly Gly Val Asp Ser Ser Pro Gln Gly
 65 70 75 80
 Asp Thr Ser Pro Ser Glu Gly Thr Asn Leu Asn Ala Pro Asn Ser Leu
 85 90 95
 Gly Val Ser Ala Leu Cys Ala Ile Cys Gly Asp Arg Ala Thr Gly Lys
 100 105 110
 His Tyr Gly Ala Ser Ser Cys Asp Gly Cys Lys Gly Phe Phe Arg Arg
 115 120 125
 Ser Val Arg Lys Asn His Met Tyr Ser Cys Arg Phe Ser Arg Gln Cys
 130 135 140
 Val Val Asp Lys Asp Lys Arg Asn Gln Cys Arg Tyr Cys Arg Leu Lys
 145 150 155 160

	Lys	Cys	Phe	Arg	Ala	Gly	Met	Lys	Lys	Glu	Ala	Val	Gln	Asn	Glu	Arg	
					165					170					175		
5	Asp	Arg	Ile	Ser	Thr	Arg	Arg	Ser	Ser	Tyr	Glu	Asp	Ser	Ser	Leu	Phe	
				180					185					190			
	Ser	Ile	Asn	Ala	Leu	Leu	Gln	Ala	Glu	Val	Leu	Ser	Arg	Gln	Ile	Thr	
10			195					200					205				
	Ser	Pro	Val	Ser	Gly	Ile	Asn	Gly	Asp	Ile	Arg	Ala	Lys	Lys	Ile	Ala	
		210					215					220					
15	Ser	Ile	Ala	Asp	Val	Cys	Glu	Ser	Met	Lys	Glu	Gln	Leu	Leu	Val	Leu	
		225				230					235					240	
	Val	Glu	Trp	Ala	Lys	Tyr	Ile	Pro	Ala	Phe	Cys	Glu	Leu	Pro	Leu	Asp	
				245						250					255		
20	Asp	Gln	Val	Ala	Leu	Leu	Arg	Ala	His	Ala	Gly	Glu	His	Leu	Leu	Leu	
				260					265					270			
	Gly	Ala	Thr	Lys	Arg	Ser	Met	Val	Phe	Lys	Asp	Val	Leu	Leu	Leu	Gly	
25			275					280					285				
	Asn	Asp	Tyr	Ile	Val	Pro	Arg	His	Cys	Pro	Glu	Leu	Ala	Glu	Met	Ser	
		290					295					300					
30	Arg	Val	Ser	Ile	Arg	Ile	Leu	Asp	Glu	Leu	Val	Leu	Pro	Phe	Gln	Glu	
		305				310					315					320	
	Leu	Gln	Ile	Asp	Asp	Asn	Glu	Tyr	Ala	Tyr	Leu	Lys	Ala	Ile	Ile	Phe	
				325						330					335		
35	Phe	Asp	Pro	Asp	Ala	Lys	Gly	Leu	Ser	Asp	Pro	Gly	Lys	Ile	Lys	Arg	
				340					345					350			
	Leu	Arg	Ser	Gln	Val	Gln	Val	Ser	Leu	Glu	Asp	Tyr	Ile	Asn	Asp	Arg	
40			355					360					365				
	Gln	Tyr	Asp	Ser	Arg	Gly	Arg	Phe	Gly	Glu	Leu	Leu	Leu	Leu	Leu	Pro	
		370					375					380					
45	Thr	Leu	Glu	Ser	Ile	Thr	Trp	Gln	Met	Ile	Glu	Gln	Ile	Gln	Phe	Ile	
		385				390					395					400	
	Lys	Leu	Phe	Gly	Met	Ala	Lys	Ile	Asp	Asn	Leu	Leu	Gln	Glu	Met	Leu	
				405						410					415		
50	Leu	Gly	Gly	Ser	Pro	Ser	Asp	Ala	Pro	His	Ala	His	His	Pro	Leu	His	
				420					425					430			
	Pro	His	Leu	Met	Gln	Glu	His	Met	Gly	Thr	Asn	Val	Ile	Val	Ala	Asn	
			435					440					445				
55	Thr	Met	Pro	Thr	His	Leu	Ser	Asn	Gly	Gln	Met	Cys	Glu	Trp	Pro	Arg	
		450					455					460					
60	Pro	Arg	Gly	Gln	Ala	Ala	Thr	Pro	Glu	Thr	Pro	Gln	Pro	Ser	Pro	Pro	
		465				470					475					480	

Gly Gly Ser Gly Ser Glu Pro Tyr Lys Leu Leu Pro Gly Ala Val Ala
485 490 495

Thr Ile Val Lys Pro Leu Ser Ala Ile Pro Gln Pro Thr Ile Thr Lys
500 505 510

Gln Glu Val Ile
515

(2) INFORMATION FOR SEQ ID NO:141:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:141:

GCGGGACCGG ATCAGCA

17

(2) INFORMATION FOR SEQ ID NO:142:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 5 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:142:

Arg Asp Arg Ile Ser
1 5

(2) INFORMATION FOR SEQ ID NO:143:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:143:

GCGGGACTGG ATCAGCA

17

(2) INFORMATION FOR SEQ ID NO:144:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 7 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:144:

Ala Glu Val Leu Ser Arg Gln
1 5

(2) INFORMATION FOR SEQ ID NO:145:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 30 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 16
(D) OTHER INFORMATION: /note= "N = C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:145:

GCGGAGGTCC TGTCNGACA GGTACCGGGG

30

(2) INFORMATION FOR SEQ ID NO:146:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 15 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: modified_base
(B) LOCATION: 8
(D) OTHER INFORMATION: /note= "N = C or T"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:146:

AAAGCAANGA GAGAT

15

(2) INFORMATION FOR SEQ ID NO:147:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 4 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ix) FEATURE:
(A) NAME/KEY: Modified-site
(B) LOCATION: 3
(D) OTHER INFORMATION: /note= "X = R or any amino acid"

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:147:

Lys Gln Xaa Glu
1